



(19)

Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 0 816 499 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
07.01.1998 Bulletin 1998/02

(51) Int Cl. 6: C12N 15/19, C07K 14/52,
C12N 5/10, C07K 1/22

(21) Application number: 97304616.2

(22) Date of filing: 27.06.1997

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC
NL PT SE

- Torigoe, Kakuji
Kurashiki-shi, Okayama (JP)
- Kurimoto, Masashi
Okayama-shi, Okayama (JP)

(30) Priority: 27.06.1996 JP 185305/96

(71) Applicant: KABUSHIKI KAISHA HAYASHIBARA
SEIBUTSU KAGAKU KENKYUJO
Okayama-shi Okayama (JP)

(74) Representative: Daniels, Jeffrey Nicholas et al
Page White & Farrer
54 Doughty Street
London WC1N 2LS (GB)

(72) Inventors:
• Okura, Takanori
Kurashiki-shi, Okayama (JP)

(54) **Genomic DNA encoding a polypeptide capable of inducing the production of interferon-gamma**

(57) Disclosed is a genomic DNA encoding a polypeptide capable of inducing the production of interferon- γ by immunocompetent cells. The genomic DNA efficiently expresses the polypeptide with high biological activities of such as inducing the production of interferon- γ by immunocompetent cells, enhancing killer cells'

cytotoxicity and inducing killer cells' formation, when introduced into mammalian host cells. The high biological activities of the polypeptide facilitate its uses to treat and/or prevent malignant tumors, viral diseases, bacterial infectious diseases and immune diseases without serious side effects when administered to humans.

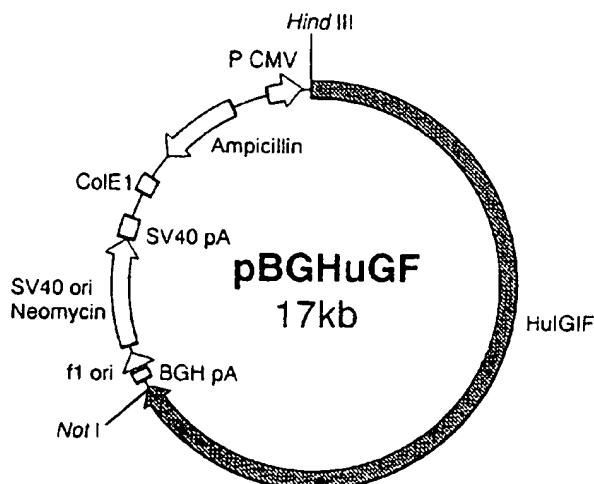


FIG.1

Description

The present invention relates to a genomic DNA, more particularly, a genomic DNA encoding a polypeptide capable of inducing the production of interferon- γ (hereinafter abbreviated as "IFN- γ ") by immunocompetent cells.

5 The present inventors successfully isolated a polypeptide capable of inducing the production of IFN- γ by immunocompetent cells and cloned a cDNA encoding the polypeptide, which is disclosed in Japanese Patent Kokai No. 27,189/96 and 193,098/96. Because the present polypeptide possesses the properties of enhancing killer cells' cytotoxicity and inducing killer cells' formation as well as inducing IFN- γ , a useful biologically active protein, it is expected to be widely used as an agent for viral diseases, microbial diseases, tumors and/or immunopathies, etc.

10 It is said that a polypeptide generated by a gene expression may be partially cleaved and/or glycosylated by processing with intracellular enzymes in human cells. A polypeptide to be used in therapeutic agents should be preferably processed similarly as in human cells, whereas human cell lines generally have a disadvantage of less producing the present polypeptide, as described in Japanese Patent Application No.269,105/96. Therefore, recombinant DNA techniques should be applied to obtain the present polypeptide in a desired amount. To produce the polypeptide processed similarly as in human cells using recombinant DNA techniques, mammalian cells should be used as the hosts.

15 In view of foregoing, the first object of the present invention is to provide a DNA which efficiently expresses the polypeptide production when introduced into a mammalian host cell.

The second object of the present invention is to provide a transformant into which the DNA is introduced.

The third object of the present invention is to provide a process for preparing a polypeptide, using the transformant.

[Means to Attain the Object]

20 The present inventors' energetic studies to attain the above objects succeeded in the finding that a genomic DNA encoding the present polypeptide efficiently expresses the polypeptide production when introduced into mammalian host cells. They found that the polypeptide thus obtained possessed significantly higher biological activities than that obtained by expressing a cDNA encoding the polypeptide in *Escherichia coli*.

25 The first object of the present invention is attained by a genomic DNA encoding a polypeptide with the amino acid sequence of SEQ ID NO:1 (where the symbol "Xaa" means "isoleucine" or "threonine") or its homologous one, which induces interferon- γ production by immunocompetent cells.

30 The second object of the present invention is attained by a transformant formed by introducing the genomic DNA into a mammalian host cell.

The third object of the present invention is attained by a process for preparing a polypeptide, which comprises (a) culturing the transformant in a nutrient medium, and (b) collecting the polypeptide from the resultant culture.

FIG.1 is a restriction map of a recombinant DNA containing a genomic DNA according to the present invention.

35 Explanation of the symbols are as follows: The symbol "*Hin* dIII" indicates a cleavage site by a restriction enzyme *Hin* dIII, and the symbol "*HuGIF*" indicates a genomic DNA according to the present invention.

The followings are the preferred embodiments according to the present invention. This invention is made based on the identification of a genomic DNA encoding the polypeptide with the amino acid sequence of SEQ ID NO:1 or its homologous one, and the finding that the genomic DNA efficiently expresses the polypeptide with high biological activities when introduced into mammalian host cells. The genomic DNA of the present invention usually contains two or more exons, at least one of which possesses a part of or the whole of the nucleotide sequence of SEQ ID NO:2. The wording "a part" includes a nucleotide and a sequential nucleotides consisting of two or more nucleotides in SEQ ID NO:2. Examples of the exons are SEQ ID NOs:3 and 4. Human genomic DNA may contain additional exons with SEQ ID NOs:5 to 7. Since the present genomic DNA is derived from a mammalian genomic DNA, it contains introns, as a distinctive feature in mammalian genomic DNAs. The present genomic DNA usually has two or more introns such as SEQ ID NOs:8 to 12.

40 More particular examples of the present genomic DNA include DNAs with SEQ ID NOs:13 and 14 or complementary sequences thereunto. The DNAs with SEQ ID NOs:13 and 14 are substantially the same. The DNA with SEQ ID NO:14 contains coding regions for a leader peptide, consisting of the nucleotides 15,607th-15,685th, 17,057th-17,068th and 20,452nd-20,468th, coding regions for the present polypeptide, consisting of the nucleotides 20,469th-20,586th, 21,921st-22,054th and 26,828th-27,046th, and regions as introns, consisting of the nucleotides 15,686th-17,056th, 17,069-20,451st, 20,587th-21,920th and 22,055th-26,827th. The genomic DNA with SEQ ID NO:13 is suitable for expressing the polypeptide in mammalian host cells.

45 Generally in this field, when artificially expressing a DNA encoding a polypeptide in a host, one or more nucleotides in a DNA may be replaced by different ones, and appropriate promoter(s) and/or enhancer(s) may be linked to the DNA to improve the expressing efficiency or the properties of the expressed polypeptide. The present genomic DNA can be altered similarly as above. Therefore, as far as not substantially changing in the biological activities of the expressed polypeptides, the present genomic DNA should include DNAs encoding functional equivalents of the

polypeptide, formed as follows: One or more nucleotides in SEQ ID NOs:3 to 14 are replaced by different ones, the untranslated regions and/or the coding region for a leader peptide in the 5'- and/or 3'-termini of SEQ ID NOs:3, 4, 5, 6, 7, 13 and 14 are deleted, and appropriate oligonucleotides are linked to either or both ends of SEQ ID NO:13.

The present genomic DNA includes general DNAs which are derived from a genome containing the nucleotide sequences as above, and it is not restricted to its sources or origins as far as it is once isolated from its original organisms. For example, the present genomic DNA can be obtained by chemically synthesizing based on SEQ ID NOs:2 to 14, or by isolating from a human genomic DNA. The isolation of the present genomic DNA from such a human genomic DNA comprises (a) isolating a genomic DNA from human cells by conventional methods, (b) screening the genomic DNA with probes or primers, which are chemically synthesized oligonucleotides with a part of or the whole of the nucleotide sequence of SEQ ID NO:2, and (c) collecting a DNA to which the probes or primers specifically hybridize. Once the present genomic DNA is obtained, it can be unlimitedly replicated by constructing a recombinant DNA with an autonomously replicable vector by conventional method and then introducing the recombinant DNA into an appropriate host such as a microorganism or an animal cell before culturing the transformant or by applying a PCR method.

The present genomic DNA is very useful in producing the polypeptide by recombinant DNA techniques since it efficiently expresses the polypeptide with high biological activities when introduced into mammalian host cells. The present invention further provides a process for preparing a polypeptide using a specific genomic DNA, comprising the steps of (a) culturing a transformant formed by introducing the present genomic DNA into mammalian host cells, and (b) collecting the polypeptide which induces IFN- γ production by immunocompetent cells from the resultant culture.

The following explains the process for preparing the polypeptide according to the present invention. The present genomic DNA is usually introduced into host cells in the form of a recombinant DNA. The recombinant DNA, comprising the present genomic DNA and an autonomously replicable vector, can be relatively easily prepared by conventional recombinant DNA techniques when the genomic DNA is available. The vectors, into which the present genomic DNA can be inserted, include plasmid vectors such as pcD, pcDL-SR α , pKY4, pCDM8, pCEV4 and pME18S. The autonomously replicable vectors usually further contain appropriate nucleotide sequences for the expression of the present recombinant DNA in each host cell, which include sequences for promoters, enhancers, replication origins, transcription termination sites, splicing sequences and/or selective markers. Heat shock protein promoters or IFN- α promoters, as disclosed in Japanese Patent Kokai No.163,368/95 by the same applicant of this invention, enables to artificially regulate the present genomic DNA expression by external stimuli.

To insert the present genomic DNA into vectors, conventional methods used in this field can be arbitrarily used: Genes containing the present genomic DNA and autonomously replicable vectors are cleaved with restriction enzymes and/or ultrasonic, and the resultant DNA fragments and the resultant vector fragments are ligated. To cleave genes and vectors by restriction enzymes, which specifically act on nucleotides, more particularly, *Acc*I, *Bam*HI, *Bgl*II, *Bst*XI, *Eco*RI, *Hind*III, *Not*I, *Pst*I, *Sac*I, *Sal*I, *Smal*, *Spel*, *Xba*I, *Xho*I, etc., facilitate the ligation of the DNA fragments and the vector fragments. To ligate the DNA fragments and the vector fragments, they are, if necessary, first annealed, then treated with a DNA ligase *in vivo* or *in vitro*. The recombinant DNAs thus obtained can be unlimitedly replicated in hosts derived from microorganisms or animals.

Any cells conventionally used as hosts in this field can be used as the host cells: Examples of such are epithelial, interstitial and hemopoietic cells, derived from human, monkey, mouse and hamster, more particularly, 3T3 cells, C127 cells, CHO cells, CV-1 cells, COS cells, HeLa cells, MOP cells and their mutants. Cells which inherently produce the present polypeptide also can be used as the host cells: Example of such are human hemopoietic cells such as lymphoblasts, lymphocytes, monoblasts, monocytes, myeloblasts, myelocytes, granulocytes and macrophages, and human epithelial and interstitial cells derived from solid tumors such as pulmonary carcinoma, large bowel cancer and colon cancer. More particular examples of the latter hemopoietic cells are leukemia cell lines such as HBL-38 cells, HL-60 cells ATCC CCL240, K-562 cells ATCC CCL243, KG-1 cells ATCC CCL246, Mo cells ATCC CRL8066, THP-1 cells ATCC TIB202, U-937 cells ATCC CRL1593.2, described by J. Minowada et al. in "Cancer Research", Vol. 10, pp. 1-18 (1988), derived from leukemias or lymphoma including myelogenous leukemias, promyelocytic leukemias, monocytic leukemias, adult T-cell leukemias and hairy cell leukemias, and their mutants. The present polypeptide-processibility of these leukemia cell lines and their mutants is so distinguished that they can easily yield the polypeptide with higher biological activities when used as hosts.

To introduce the present DNA into the hosts, conventional methods such as DEAE-dextran method, calcium phosphate transfection method, electroporation method, lipofection method, microinjection method, and viral infection method as using retrovirus, adenovirus, herpesvirus and vaccinia virus, can be used. The polypeptide-producing clones in the transformants can be selected by applying the colony hybridization method or by observing the polypeptide production after culturing the transformants in culture media. For example, the recombinant DNA techniques using mammalian cells as hosts are detailed in "Jikken-Igaku-Bessatsu Saibo-Kogaku Handbook (The handbook for the cell engineering)" (1992), edited by Toshio KUROKI, Masaru TANIGUCHI and Mitsuo OSHIMURA, published by YODOSHA CO., LTD., Tokyo, Japan, and "Jikken-Igaku-Bessatsu Biomanual Series 3 Idenshi Cloning Jikken-Ho (The experimen-

tal methods for the gene cloning)" (1993), edited by Takahi YOKOTA and Ken-ichi ARAI, published by YODOSHA CO., LTD., Tokyo, Japan.

The transformants thus obtained secrete the present polypeptide intracellularly and/or extracellularly when cultured in culture media. As the culture media, conventional ones used for mammalian cells can be used. The culture media generally comprise (a) buffers as a base, (b) inorganic ions such as sodium ion, potassium ion, calcium ion, phosphoric ion and chloric ion, (c) micronutrients, carbon sources, nitrogen sources, amino acids and vitamins, which are added depending on the metabolic ability of the cells, and (d) sera, hormones, cell growth factors and cell adhesion factors, which are added if necessary. Examples of individual media include 199 medium, DMEM medium, Ham's F12 medium, IMDM medium, MCDB 104 medium, MCDB 153 medium, MEM medium, RD medium, RITC 80-7 medium, RPMI-1630 medium, RPMI-1640 medium and WAJC 404 medium. The cultures containing the present polypeptide are obtainable by inoculating the transformants into the culture media to give a cell density of 1×10^4 - 1×10^7 cells/ml, more preferably, 1×10^5 - 1×10^6 cells/ml, and then subjecting to suspension- or monolayer-cultures at about 37°C for 1-7 days, more preferably, 2-4 days, while appropriately replacing the culture media with a fresh preparation of the culture media. The cultures thus obtained usually contain the present polypeptide in a concentration of about 1-100 µg/ml, which may vary depending on the types of the transformants or the culture conditions used.

While the cultures thus obtained can be used intact as an IFN-γ inducer, they are usually subjected to a step for separating the present polypeptide from the cells or the cell debris using filtration, centrifugation, etc. before use, which may follow a step for disrupting the cells with supersonication, cell-lytic enzymes and/or detergents if desired, and to a step for purifying the polypeptide. The cultures from which the cells or cell debris are removed are usually subjected to conventional methods used in this field for purifying biologically active polypeptides, such as salting-out, dialysis, filtration, concentration, separatory sedimentation, ion-exchange chromatography, gel filtration chromatography, adsorption chromatography, chromatofocusing, hydrophobic chromatography, reversed phase chromatography, affinity chromatography, gel electrophoresis and/or isoelectric focusing. The resultant purified polypeptide can be concentrated and/or lyophilized into liquids or solids depending on final uses. The monoclonal antibodies disclosed in Japanese Patent Kokai No.231,598/96 by the same applicant of this invention are extremely useful to purify the present polypeptide. Immunoaffinity chromatography using monoclonal antibodies yields the present polypeptide in a relatively high purity at the lowest costs and labors.

The polypeptide obtainable by the process according to the present invention exerts strong effects in the treatment and/or the prevention for IFN-γ- and/or killer cell-susceptive diseases since it possesses the properties of enhancing killer cells' cytotoxicity and inducing killer cells' formation as well as inducing IFN-γ, a useful biologically active protein, as described above. The polypeptide according to the present invention has a high activity of inducing IFN-γ, and this enables a desired amount of IFN-γ production with only a small amount. The polypeptide is so low toxic that it scarcely causes serious side effects even when administered in a relatively-high dose. Therefore, the polypeptide has an advantage that it can readily induce IFN-γ in a desired amount without strictly controlling the dosage. The uses as agents for susceptive diseases are detailed in Japanese Patent Application No.28,722/96 by the same applicant of this invention.

The present genomic DNA is also useful for so-called "gene therapy". According to conventional gene therapy, the present DNA can be introduced into patients with IFN-γ- and/or killer cell-susceptive diseases by directly injecting after the DNA is inserted into vectors derived from viruses such as retrovirus, adenovirus and adeno-associated virus or is incorporated into cationic- or membrane fusible-liposomes, or by self-transplanting lymphocytes which are collected from patients before the DNA is introduced. In adoptive immunotherapy with gene therapy, the present DNA is introduced into effector cells similarly as in conventional gene therapy. This can enhance the cytotoxicity of the effector cells to tumor cells, resulting in improvement of the adoptive immunotherapy. In tumor vaccine therapy with gene therapy, tumor cells from patients, into which the present genomic DNA is introduced similarly as in conventional gene therapy, are self-transplanted after proliferated *ex vivo* up to give a desired cell number. The transplanted tumor cells act as vaccines in the patients to exert a strong antitumor immunity specifically to antigens. Thus, the present genomic DNA exhibits considerable effects in gene therapy for diseases including viral diseases, microbial diseases, malignant tumors and immunopathies. The general procedures for gene therapy are detailed in "*Jikken-Igaku-Bessatsu Biomanual UP Series Idenshichiryo-no-Kisogijutsu* (Basic techniques for the gene therapy)" (1996), edited by Takashi ODA-JIMA, Izumi SAITO and Keiya OZAWA, published by YODOSHA CO., LTD., Tokyo, Japan.

The following examples explain the present invention, and the techniques used therein are conventional ones used in this field: For example, the techniques are described in "*Jikken-Igaku-Bessatsu Saibo-Kogaku Handbook* (The handbook for the cell engineering)", (1992), edited by Toshio KUROKI, Masaru TANIGUCHI and Mitsuo OSHIMURA, published by YODOSHA CO., LTD., Tokyo, Japan, and "*Jikken-Igaku-Bessatsu Biomanual Series 3 Idenshi Cloning Jikken-Ho* (The experimental methods for the gene cloning)" (1993), edited by Takahi YOKOTA and Ken-ichi ARAI, published by YODOSHA CO., LTD., Tokyo, Japan.

Example 1Cloning genomic DNA and determination of nucleotide sequence5 Example 1-1Determination of partial nucleotide sequence

10 Five ng of "PromoterFinder™ DNA Pvull LIBRARY", a human placental genomic DNA library commercialized by CLONTECH Laboratories, Inc., California, USA, 5 μ l of 10 x Tth PCR reaction solution, 2.2 μ l of 25 mM magnesium acetate, 4 μ l of 2.5 mM dNTP-mixed solution, one μ l of the mixed solution of 2 unit/ μ l rTth DNA polymerase XL and 2.2 μ g/ μ l Tth Start Antibody in a ratio of 4:1 by volume, 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-CCATCCTAATACGACTCACTATAGGGC-3' as an adaptor primer, and 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-TTCCTCTTCCGAAGCTGTAGACTGC-3' as an anti-sense primer, which was chemically synthesized based on the sequence of the nucleotides 88th-115th in SEQ ID NO:2, were mixed and volumed up to 50 μ l with sterilized distilled water. After incubating at 94°C for one min, the mixture was subjected to 7 cycles of incubations at 94°C for 25 sec and at 72°C for 4 min, followed by 32 cycles of incubations at 94°C for 25 sec at 67°C for 4 min to perform PCR.

20 The reaction mixture was diluted by 100 folds with sterilized distilled water. One μ l of the dilution, 5 μ l of 10 x Tth PCR reaction solution, 2.2 μ l of 25 mM magnesium acetate, 4 μ l of 2.5 mM dNTP-mixed solution, one μ l of the mixed solution of 2 unit/ μ l rTth DNA polymerase XL and 2.2 μ g/ μ l Tth Start Antibody in a ratio of 4:1 by volume, 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-CTATAGGGCACGCGTGGT-3' as a nested primer, and 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-TTCCTCTTCCGAAGCTGTAGACTGC-3' as an anti-sense primer, which was chemically synthesized similarly as above, were mixed and volumed up to 50 μ l with sterilized distilled water. After incubating at 94°C for one min, the mixture was subjected to 5 cycles of incubations at 94°C for 25 sec and at 72°C for 4 min, followed by 22 cycles of incubations at 94°C for 25 sec and at 67°C for 4 min to perform PCR for amplifying a DNA fragment of the present genomic DNA. The genomic DNA library and reagents for PCR used above were mainly from "PromoterFinder™ DNA WALKING KITS", commercialized by CLONTECH Laboratories, Inc., California, USA

25 30 An adequate amount of the PCR product thus obtained was mixed with 50 ng of "pT7 Blue(R)", a plasmid vector commercialized by Novagen, Inc., WI, USA, and an adequate amount of T4 DNA ligase, and 100 mM ATP was added to give a final concentration of one mM, followed by incubating at 16°C for 18 hr to insert the DNA fragment into the plasmid vector. The obtained recombinant DNA was introduced into an *Escherichia coli* JM109 strain by the competent cell method to form a transformant, which was then inoculated into L-broth medium (pH 7.2) containing 50 μ g/ml ampicillin and cultured at 37°C for 18 hr. The cells were isolated from the resulting culture, and then subjected to the conventional alkali-SDS method to collect a recombinant DNA. The dideoxy method analysis confirmed that the recombinant DNA contained the DNA fragment with a sequence of the nucleotides 5,150th-6,709th in SEQ ID NO:14.

40 Example 1-2Determination of partial nucleotide sequence

45 PCR was performed in the same conditions as the first PCR in Example 1-1, but an oligonucleotide with the nucleotide sequence of 5'-GTAAGTTTCACCTTCCACTGTAGACTGC-3', which was chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-1, was used as an anti-sense primer.

50 The reaction mixture was diluted by 100 folds with sterilized distilled water. One μ l of the dilution was placed into a reaction tube, and PCR was performed in the same conditions as used in the second PCR in Example 1-1 to amplify another DNA fragment of the present genomic DNA, but an oligonucleotide with the nucleotide sequence of 5'-GGGAT-CAAGTAGTGATCAGAACGACAC-3', which was chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-1, was used as an anti-sense primer.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 1st-5,228th in SEQ ID NO:14.

Example 1-3Determination of partial nucleotide sequence

5 0.5 µg of a human placental genomic DNA, commercialized by CLONTECH Laboratories, Inc., California, USA, 5 µl of 10 x PCR reaction solution, 8 µl of 2.5 mM dNTP-mixed solution, one µl of the mixed solution of 5 unit/µl "TAKARA LA Taq POLYMERASE" and 1.1 µg/µl "TaqStart ANTIBODY" in a ratio of 1:1 by volume, both of them are commercialized by Takara Syuzo Co., Tokyo, Japan, 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-CCTGGCT-GCCAACCTGGCTGCTAAAGCGG-3' as a sense primer, chemically synthesized based on a sequence of the nucleotides 46th-75th in SEQ ID NO:2, and 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-GTATTGT-CAATAAATTTCATTGCCACAAAGTTG-3' as an anti-sense primer, chemically synthesized based on a sequence of the nucleotides 210th-242nd in SEQ ID NO:2, were mixed and volumed up to 50 µl with sterilized distilled water. After incubating at 94°C for one min, the mixture was subjected to 5 cycles of incubations at 98°C for 20 sec and at 68°C for 10 min, followed by 25 cycles of incubations at 98°C for 20 sec and 68°C for 10 min, with adding 5 sec in times to every cycle, and finally incubated at 72°C for 10 min to amplify further DNA fragment of the present genomic DNA. The reagents for PCR used above were mainly from "TAKARA LA PCR KIT VERSION 2", commercialized by Takara Syuzo Co., Tokyo, Japan.

10 The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 6,640th-15,671st in SEQ ID NO:14.

Experiment 1-4Determination of partial nucleotide sequence

25 PCR was performed in the same conditions as the PCR in Example 1-3 to amplify further another DNA fragment of the present genomic DNA; but an oligonucleotide with the nucleotide sequence of 5'-AAGATGGCTGCTGAACCAG-TAGAAGACAATTGC-3', chemically synthesized based on a sequence of the nucleotide 175th-207th in SEQ ID NO:2, was used as a sense primer, an oligonucleotide with the nucleotide sequence of 5'-TCCTTGGTCAATGAAGA-GAACTTGGTC-3', chemically synthesized based on a sequence of nucleotides 334th-360th in the SEQ ID NO:2, was used as an anti-sense primer, and after incubating at 98°C for 20 sec, the reaction mixture was subjected to 30 cycles of incubations at 98°C for 20 sec and at 68°C for 3 min, followed by incubating at 72°C for 10 min.

30 The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 15,604th-20,543rd in SEQ ID NO:14.

Example 1-5Determination of partial nucleotide sequence

40 PCR was performed in the same conditions as the PCR in Example 1-4 to amplify further another DNA fragment of the present genomic DNA, but an oligonucleotide with the nucleotide sequence of 5'-CCTGGAATCAGATTACTTT-GGCAAGCTTGAATC-3', chemically synthesized based on the sequence of the nucleotide 273rd-305th in SEQ ID NO:2, was used as a sense primer, and an oligonucleotide with the nucleotide sequence of 5'-GGAAATAATTTGTTCT-CACAGGAGAGAGTTG-3', chemically synthesized based on the sequence of nucleotides 500th-531st in the SEQ ID NO:2, was used as an anti-sense primer.

45 The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 20,456th-22,048th in SEQ ID NO:14.

Example 1-6Determination of partial nucleotide sequence

55 PCR was performed in the same conditions as the PCR in Example 1-4 to amplify further another DNA fragment

of the present genomic DNA, but an oligonucleotide with the nucleotide sequence of 5'-GCCAGCCTAGAGGTATGGCT-GTAACTATCTC-3', chemically synthesized based on the sequence of the nucleotide 449th-479th in SEQ ID NO:2, was used as a sense primer, and an oligonucleotide with the nucleotide sequence of 5'-GGCATGAAATTAAAT-AGCTAGTCTCGTTTG-3', chemically synthesized based on the sequence of nucleotides 745th-777th in the SEQ ID NO:2, was used as an anti-sense primer.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 21,996th-27,067th in SEQ ID NO:14.

Example 1-7

Determination of partial nucleotide sequence

PCR was performed in the same conditions as the first PCR in Example 1-2 to amplify further another DNA fragment in the present genomic DNA, but an oligonucleotide with the nucleotide sequence of 5'-GTGACATCATATTCTTCA-GAGAAGTGTCC-3', chemically synthesized based on the sequence of the nucleotide 575th-604th in SEQ ID NO:2, was used as a sense primer.

The reaction mixture was diluted by 100 folds with sterilized distilled water. One μ l of the dilution was placed into a reaction tube, and PCR was performed in the same conditions as the second PCR in Example 1-2 to amplify further another DNA fragment of the present genomic DNA, but an oligonucleotide with the sequence of 5'-GCAATTGAATCT-TCATCATACGAAGGATAC-3', chemically synthesized based on a sequence of the nucleotides 624th-654th in SEQ ID NO:2, was used as a sense primer.

The DNA fragment was inserted into the plasmid vector similarly as in Example 1-1 to obtain a recombinant DNA. The recombinant DNA was replicated in *Escherichia coli* before being collected. The analysis of the collected recombinant DNA confirmed that it contained the DNA fragment with a sequence of the nucleotides 26,914th-28,994th in SEQ ID NO:14.

Example 1-8

Determination of complete nucleotide sequence

Comparing the nucleotide sequence of SEQ ID NO:2, which was proved to encode the present polypeptide, as disclosed in Japanese Patent Kokai No.193,098/96 by the same applicant of this invention, with the partial nucleotide sequences identified in Examples 1-1 to 1-7, it was proved that the present genomic DNA contained the nucleotide sequence of SEQ ID NO:14. SEQ ID NO:14, consisting of 28,994 base pairs (bp), was extremely longer than the SEQ ID NO:2, consisting of only 471 bp. This suggested that SEQ ID NO:14 contained introns, a characteristic of eukalytic cells.

It was examined where partial nucleotide sequences of SEQ ID NO:2, i.e., exons, and the donor and acceptor sites in introns, respectively consisting of the nucleotides of GT and AG, located in SEQ ID NO:14. Consequently, it was proved that SEQ ID NO:14 contained at least 5 introns, which located in the order of SEQ ID NOs:10, 11, 12, 8 and 9 in the direction from the 5'- to the 3'-termini. Therefore, the sequences between the neighboring introns must be exons, which were thought to be located in the order of SEQ ID NOs:5, 6, 3, 4 and 7 in the direction from the 5'- to the 3'-termini. It was also proved that SEQ ID NO:7 contained the 3'-untranslated region other than the exons. The features of the sequence elucidated as this are arranged in SEQ ID NO:14.

As disclosed in Japanese Patent Application No. 269,105/96 by the same applicant of this invention, the present polypeptide is produced as a polypeptide with N-terminal amino acid of tyrosine other than methionine in human cells, which is observed in SEQ ID NO:1. This suggests that the present genomic DNA contains a leader peptide region in the upstream of the 5'-terminus of the present polypeptide-encoding region. A sequence consisting of 36 amino acids encoded by the upstream of the nucleotides 20,469th-20,471st, which is the nucleotides of TAC, are described as a leader peptide in SEQ ID NO:14.

Example 2

Preparation of recombinant DNA pBGHuGF for expression

0.06 ng of the DNA fragment in Example 1-4 in a concentration of 3 ng/50 μ l, 0.02 ng of the DNA fragment, obtained by the methods in Example 1-5, 5 μ l of 10 x LA PCR reaction solution, 8 μ l of 2.5 mM dNTP-mixed solution, one μ l of

the mixed solution of 5 unit/ μ l TAKARA LA Taq polymerase and 1.1 μ g/ μ l TaqStart Antibody in a ratio of 1:1 by volume, 10 pmol of an oligonucleotide with the sequence of 5'-TCGAAAGCTTAAGATGGCTGCTGAACCAGTA-3' as a sense primer, chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-4, and 10 pmol of an oligonucleotide with the nucleotide sequence of 5'-GGAAATAATTTGTTCTCACAGGAGAGAGTTG-3' as an anti-sense primer, chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-5, were mixed and volumed up to 50 μ l with sterilized distilled water. After incubating at 94°C for one min, the mixture was subjected to 5 cycles of incubations at 98°C for 20 sec and at 72°C for 7 min, followed by 25 cycles of incubations at 98°C for 20 sec and 68°C for 7 min to perform PCR. The reaction mixture was cleaved by restriction enzymes *Hind*III and *Sph*I to obtain a DNA fragment of about 5,900 bp, with cleavage sites by *Hind*III and *Sph*I in its both termini.

PCR was performed in the same condition as above, but 0.02 ng of the DNA fragment in Example 1-5, 0.06 ng of the DNA fragment obtained in Example 1-6, an oligonucleotide with the nucleotide sequence of 5'-ATGTAGCG-GCCGCGGCATGAAATTTAATAGCTAGTC-3' as an anti-sense primer, chemically synthesized based on the nucleotide sequence of the DNA fragment in Example 1-6, and an oligonucleotide with the sequence of 5'-CCTGGAATCA-GATTACTTGGCAAGCTTGAATC-3' as a sense primer, chemically synthesized based on the DNA fragment in Example 1-6, were used. The reaction mixture was cleaved by restriction enzymes *Nol*I and *Sph*I to obtain a DNA fragment of about 5,600 bp, with cleavage sites by *Nol*I and *Sph*I in its both termini.

A plasmid vector "pRc/CMV", containing a *cytomegalovirus* promoter, commercialized by Invitrogen Corporation, San Diego, USA, was cleaved by restriction enzymes *Hind*III and *Nol*I to obtain a vector fragment of about 5,500 bp. The vector fragment was mixed with the above two DNA fragments of about 5,900 bp and 5,600 bp, and reacted with T4 DNA ligase to insert the two DNA fragments into the plasmid vector. An *Escherichia coli* JM109 strain was transformed with the obtained recombinant DNA, and the transformant with the plasmid vector was selected by the colony hybridization method. The selected recombinant DNA was named as "pBGHuGF". As shown in FIG.1, the present genomic DNA, with the nucleotide sequence of SEQ ID NO:13, was ligated in the downstream of the cleavage site by the restriction enzyme *Hind*III in the recombinant DNA.

25

Example 3

Preparation of transformant using CHO cell as host

CHO-K1 cells ATCC CCL61 were inoculated into Ham's F12 medium (pH 7.2) containing 10 v/v % bovine fetal serum and proliferated by conventional manner. The proliferated cells were collected and washed with phosphate-buffered saline (hereinafter abbreviated as "PBS") followed by suspending in PBS to give a cell density of 1 x 10⁷ cells/ml.

10 μ g of the recombinant DNA pBGHuGF in Example 2 and 0.8 ml of the above cell suspension were placed in a cuvette and ice-chilled for 10 min. The cuvette was installed in "GENE PULSER", an electroporation device commercialized by Bio-Rad Laboratories Inc., Brussels, Belgium, and then pulsed once with an electric discharge. After pulsing, the cuvette was immediately took out and ice-chilled for 10 min. The cell suspension from the cuvette was inoculated into Ham's F12 medium (pH 7.2) containing 10 v/v % bovine fetal serum and cultured under an ambient condition of 5 v/v % CO₂ at 37°C for 3 days. To the culture medium was added G-418 to give a final concentration of 400 μ g/ml, and the culturing was continued further 3 weeks under the same conditions. From about 100 colonies formed, 48 colonies were selected, and a portion of each was inoculated into a well of culturing plates with Ham's F12 medium (pH 7.2) containing 400 μ g/ml G-418 and 10 v/v % bovine fetal serum and cultured similarly as above. Thereafter, to each well of the culturing plates was added 10 mM Tris-HCl buffer (pH 8.5) containing 5.1 mM magnesium chloride, 0.5 w/v % sodium deoxycholate, 1 w/v % NONIDET P-40, 10 μ g/ml aprotinin and 0.1 w/v % SDS to lyse the cells.

50 μ l aliquot of the cell lysates was mixed with one ml of glycerol and incubated at 37°C for one hour, before the polypeptides in the cell lysates were separated by the SDS-polyacrylamide gel electrophoresis. The separated polypeptides were transferred to a nitrocellulose membrane in usual manner, and the membrane was soaked in the culture supernatant of the hybridoma H-1, disclosed in Japanese Patent Kokai No.231,598/96 by the same applicant of this invention, followed by washing with 50 mM Tris-HCl buffer containing 0.05 v/v % TWEEN 20 to remove an excessive mount of the monoclonal antibody. Thereafter, the nitrocellulose membrane was soaked in PBS containing rabbit-derived anti-mouse immunoglobulin antibody for one hr, which was labeled with horseradish peroxidase, followed by washing 50 mM Tris-HCl buffer (pH 7.5) containing 0.05 v/v % TWEEN 20 and soaking in 50 mM Tris-HCl buffer (pH 7.5) containing 0.005 v/v % hydrogen peroxide and 0.3 mg/ml diaminobenzidine to develop colorations. The clone, which highly produced the polypeptide, was selected based on the color development and named "BGHuGF".

55

Example 4Production of polypeptide by transformant and its physicochemical properties

5 The transformant BGHuGF in Experiment 3 was inoculated into Ham's F12 medium (pH 7.2) containing 400 µg/ml G-418 and 10 v/v % bovine fetal serum, and cultured under an ambient condition of 5 v/v % CO₂ at 37°C for one week. The proliferated cells were collected, washed with PBS, and then washing with 10-fold volumes of ice-chilled 20 mM Hepes buffer (pH 7.4), containing 10 mM potassium chloride and 0.1 mM ethylenediaminetetraacetate bisodium salt, according to the method described in "Proceedings of The National Academy of The Sciences of The USA", vol. 10, pp.5,227-5,231 (1989), by M. J. Kostura et al. The cells thus obtained were allowed to stand in 3-fold volumes of a fresh preparation of the same buffer under an ice-chilling condition for 20 min and freezed at -80°C, succeeded by thawing to disrupt the cells. The resulting cells were centrifuged to collect the supernatant.

10 In parallel, THP-1 cells ATCC TIB 202, derived from a human acute monocytic leukemia, was similarly cultured and disrupted. Supernatant, obtained by centrifuging the resulting cells, was mixed with the supernatant obtained from the transformant BGHuGF and incubated at 37°C for 3 hr to react. The reaction mixture was applied to a column with "DEAE-SEPHAROSE", a gel for ion-exchange chromatography, commercialized by Pharmacia LKB Biotechnology AB, Upsalla, Sweden, equilibrated with 10 mM phosphate buffer (pH 6.6) before use. After washing the column with 10 mM phosphate buffer (pH 6.6), 10 mM phosphate buffer (pH 6.6) with a stepwise gradient of NaCl increasing from 0 M to 0.5 M was fed to the column, and fractions eluted by about 0.2 M NaCl were collected. The fractions were dialyzed against 10 mM phosphate buffer (pH 6.8) before applied to a column with "DEAE 5PW", a gel for ion-exchange chromatography, commercialized by TOSOH Corporation, Tokyo, Japan. To the column was fed 10 mM phosphate buffer (pH 6.8) with a linear gradient of NaCl increasing from 0 M to 0.5 M, and fractions eluted by about 0.2-0.3 M NaCl were collected.

15 While the obtained fractions were pooled and dialyzed against PBS, a gel for immunoaffinity chromatography with the monoclonal antibody were prepared according to the method disclosed in Japanese Patent Kokai No.231,598/96 by the same applicant of this invention. After the gel were charged into a plastic column and washed with PBS, the above dialyzed solution was applied to the column. To the column was fed 100 mM glycine-HCl buffer (pH 2.5), and the eluted fractions, which contained a polypeptide capable of inducing the production of IFN-γ by immunocompetent cells, were collected. After the collected fractions were dialyzed against sterilized distilled water and concentrated with 20 a membrane filtration, the resultant was lyophilized to obtain a purified solid polypeptide in a yield of about 15 mg/l-culture.

Example for ReferenceExpression in Escherichia coli

35 As disclosed in Japanese Patent Kokai No.193,098/96, a transformant pKHuGF which was obtained by introducing a cDNA with the nucleotide sequence of SEQ ID NO:2 into Escherichia coli as a host, was inoculated into L-broth medium containing 50 µg/ml ampicillin and cultured at 37°C for 18 hr under shaking conditions. The cells were collected by centrifuging the resulting culture, and then suspended in a mixture solution (pH 7.2) of 139 mM NaCl, 7 mM NaH₂PO₄ and 3 mM Na₂HPO₄, followed by sonication to disrupt the cells. After the cell disruptants were centrifuged, the supernatant was subjected to purifying steps similarly as in Example 4-1 to obtain a purified solid polypeptide in a yield of about 5 mg/l-culture.

40 Comparing the yields of the polypeptides in Example for Reference and in Example 4-1 shows that the use of a transformant, which is formed by introducing a genomic DNA encoding the present polypeptide into a mammalian cell as a host, strongly elevates the yield of the polypeptide per culture.

Example 4-2Physicochemical property of polypeptideExample 4-2(a)Biological activity

55 Blood were collected from a healthy donor by using a syringe containing heparin, and then diluted with 2-fold volume of serum-free RPMI-1640 medium (pH 7.4). The blood was overlaid on ficoll, commercialized by Pharmacia LKB Biotechnology AB, Upsalla, Sweden, and centrifuged to obtain lymphocytes, which were then washed with RPMI-

1640 medium containing 10 v/v % bovine fetal serum before being suspended in a fresh preparation of the same medium to give a cell density of 5×10^6 cells/ml. 0.15 ml aliquots of the cell suspension was distributed into wells of micro plates with 96 wells.

5 To the wells with the cells were distributed 0.05 ml aliquots of solutions of the polypeptide in Example 4-1, diluted with RPMI-1640 medium (pH 7.4) containing 10 v/v % bovine fetal serum to give desired concentrations. 0.05 ml aliquots of fresh preparations of the same medium with 2.5 μ g/ml concanavalin A were further added to the wells, before culturing in a 5 v/v % CO₂ incubator at 37°C for 24 hr. After the cultivation, 0.1 ml of the culture supernatant was collected from each well and examined on IFN- γ by usual enzyme immunoassay. In parallel, a systems as a control using the polypeptide in Reference for that in Example 4-1 or using no polypeptide was treated similarly as above. The 10 results were in Table 1. IFN- γ in Table 1 were expressed with international units (IU), calculated based on the IFN- γ standard, Gg23-901-530, obtained from the International Institute of Health, USA

Table 1

Sample of polypeptide	IFN- γ production (IU/ml)
Example 4-2(a)	3.4×10^5
Example for Reference	1.7×10^5

15 Table 1 indicates that the lymphocytes as immunocompetent cells produce IFN- γ by the action of the present 20 polypeptide.

It is more remarkable that the polypeptide in Example 4-1 could induce IFN- γ production more than that in Example 25 for Reference. Considering this and the difference in the yields of the polypeptides, described in Example for Reference, it can be presumed: Even if DNAs could be substantially equivalent in encoding the same amino acid sequence, not only the expressing efficiencies of the DNAs may differ, but the products expressed by them may significantly differ in their biological activities as a result of post-translational modifications by intracellular enzymes, depending on types of the DNAs and their hosts; (a) one type is used a transformant formed by introducing a DNA, which is a cDNA, into a microorganisms as a host, and (b) other type is used a transformant formed by introducing the present genomic DNA into a mammalian cell as a host.

30 Example 4-2(b)

Molecular weight

35 SDS-polyacrylamide gel electrophoresis of the polypeptide in Example 4-1 in the presence of 2 w/v % dithiothreitol as a reducing agent, according to the method reported by U. K. Laemli et al., in "Nature", Vol.227, pp.680-685 (1970), exhibited a main band of a protein capable of inducing IFN- γ in a position corresponding to a molecular weight of about 18,000-19,500 daltons. The molecular weight makers used in the analysis were bovine serum albumin (67,000 daltons), ovalbumin (45,000 daltons), carbonic anhydrase (30,000 daltons), soy bean trypsin inhibitor (20,100 daltons) and α -lactoalbumin (14,000 daltons).

40 Example 4-2(c)

N-Terminal amino acid sequence

45 Conventional analysis using "MODEL 473A", a protein sequencer commercialized by Perkin-Elmer Corp., Norwalk, USA, revealed that the polypeptide in Example 4-1 had the amino acid sequence of SEQ ID NO:15 in the N-terminal region.

50 Judging collectively from this result as well as the information that SDS-polyacrylamide gel electrophoresis exhibited a main band in a position corresponding to a molecular weight of about 18,000-19,500 daltons, and that the molecular weight calculated from the amino acid sequence of SEQ ID NO:1 was 18,199 daltons, it can be concluded that the polypeptide in Example 4-1 has the amino acid sequence of SEQ ID NO:6.

55 As is described above, the present invention is made based on the identification of a genomic DNA encoding the polypeptide which induces the production of IFN- γ by immunocompetent cells. The present genomic DNA efficiently express the present polypeptide when introduced into mammalian host cells. The polypeptide features higher biological activities than that obtained by the cDNA expression in *Escherichia coli*. Therefore, the present genomic DNA is useful for the recombinant DNA techniques to prepare the polypeptide capable of inducing IFN- γ production by immunocompetent cells. The present genomic DNA is useful to gene therapy for diseases including viral diseases, bacterial-infectious diseases, malignant tumors and immunopathies.

Thus, the present invention is a significant invention which has a remarkable effect and gives a great contribution to this field.

While there has been described what is at present considered to be the preferred embodiments of the present invention, it will be understood the various modifications may be made therein, and it is intended to cover in the appended claims all such modifications as fall within the true spirits and scope of the invention.

10

15

20

25

30

35

40

45

50

55

SEQUENCE LISTING

(1) GENERAL INFORMATION:

5 (i) APPLICANT:

NAME:KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENKYUJO

10 (ii) TITLE OF INVENTION:GENOMIC DNA ENCODING A POLYPEPTIDE
CAPABLE OF INDUCING THE PRODUCTION OF INTERFERON- γ

(iii) NUMBER OF SEQUENCES:15

15 (iv) ADDRESS:

(A) ADDRESSEE:KABUSHIKI KAISHA HAYASHIBARA SEIBUTSU KAGAKU
KENKYUJO
(B) STREET:2-3, 1-CHOME, SHIMOISHII
(C) CITY:OKAYAMA
(E) COUNTRY:JAPAN
(F) POSTAL CODE (ZIP):700

20 (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE:Floppy disk
(B) COMPUTER:IBM PC compatible
(C) OPERATING SYSTEM:PC-DOS/MS-DOS

25 (vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:JP 185,305/96
(B) FILING DATE:June 27, 1996

30 (2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH:157 amino acids
(B) TYPE:amino acid
(D) TOPOLOGY:linear

(ii) MOLECULE TYPE:peptide

35 (xi) SEQUENCE DESCRIPTION:SEQ ID NO:1:

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn
1 5 10 15
Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp
20 25 30
40 Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg Thr Ile Phe Ile
35 40 45
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
50 55 60
45 Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys Glu Asn Lys Ile
65 70 75 80
Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys
85 90 95
55 Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys
100 105 110
Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu
115 120 125
50 Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu
130 135 140
Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp
145 150 155

(3) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1120 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: No

(iv) ANTI-SENSE: No

(v) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: liver

(ix) FEATURE:

- (A) NAME/KEY: 5' UTR
- (B) LOCATION: 1..177
- (C) IDENTIFICATION METHODS: E
- (A) NAME/KEY: leader peptide
- (B) LOCATION: 178..285
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: mat peptide
- (B) LOCATION: 286..756
- (C) IDENTIFICATION METHODS: S
- (A) NAME/KEY: 3' UTR
- (B) LOCATION: 757..1120
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

25	GCCTGGACAG TCAGCAAGGA ATTGTCTCCC AGTGCATTTC	GCCCTCCTGG CTGCCAACTC	60
	TGGCTGCTAA AGCGGCTGCC ACCTGCTGCA GTCTACACAG	CTTCGGGAAG AGGAAAGGAA	120
	CCTCAGACCT TCCAGATCGC TTCCCTCTCGC AACAAACTAT	TTGTCGCAGG AATAAAG	177
	ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC	TTT GTG GCA ATG	225
30	Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn	Phe Val Ala Met	
	-35 -30 -25		
	AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA	GAT GAT GAA AAC	273
	Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu	Asp Asp Glu Asn	
	-20 -15 -10 -5		
35	CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA	TTA TCA GTC ATA	321
	Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys	Leu Ser Val Ile	
	1 5 10		
	AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA	GGA AAT CGG CCT	369
	Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln	Gly Asn Arg Pro	
	15 20 25		
40	CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA GAT	AAT GCA CCC CGG	417
	Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp	Asn Ala Pro Arg	
	30 35 40		
	ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG	CCT AGA GGT ATG	465
	Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln	Pro Arg Gly Met	
	45 50 55 60		
45	GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA	AYT CTC TCC TGT	513
	Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser	Xaa Leu Ser Cys	
	65 70 75		
	GAG AAC AAA ATT ATT TCC TTT AAG GAA ATG AAT CCT	CCT GAT AAC ATC	561
	Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro	Pro Asp Asn Ile	
	80 85 90		
50	AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA	AGT GTC CCA GGA	609
	Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg	Ser Val Pro Gly	
	95 100 105		
	CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC	GAA GGA TAC TTT	657
	His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr	Glu Gly Tyr Phe	
	110 115 120		
55	CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC	ATT TTG AAA AAA	705

125 Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys
 130 135 140
 GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT GTT CAA AAC GAA 753
 5 Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu
 145 150 155
 GAC TAGCTATTAA AATTCATGC CGGGCGCAGT GGCTCACGCC TGTAATCCCA 806
 Asp
 GCCCTTTGGG AGGCTGAGGC GGGCAGATCA CCAGAGGTCA GGTGTTCAAG ACCAGCCTGA 866
 CCAACATGGT GAAACCTCAT CTCTACTAAA AATACTAAAA ATTAGCTGAG TGTAGTGACG 926
 10 CATGCCCTCA ATCCCAGCTA CTCAAGAGGC TGAGGCAGGA GAATCACTTG CACTCCGGAG 986
 GTAGAGGTTG TGGTGAGCCG AGATTGACCC ATTGCGCTCT AGCCTGGGCA ACAACAGCAA 1046
 AACTCCATCT CAAAAAAATAA AATAAAATAA TAAACAAATA AAAAATTCTAT AATGTGAAAA 1106
 AAAAAAAA AAAA 1120

15 (4) INFORMATION FOR SEQ ID NO:3:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 135 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 20 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: human
 (F) TISSUE TYPE: placenta
 (ix) FEATURE:
 (A) NAME/KEY: exon
 (B) LOCATION: 1..135
 (C) IDENTIFICATION METHODS: S
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

30 AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA 47
 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser
 5 1 5 10
 GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT 95
 Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn
 15 20 25
 35 CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA G 135
 Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp
 30 35 40

40 (5) INFORMATION FOR SEQ ID NO:4:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 134 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 45 (ii) MOLECULE TYPE: Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: human
 (F) TISSUE TYPE: placenta
 (ix) FEATURE:
 (A) NAME/KEY: exon
 (B) LOCATION: 1..134
 (C) IDENTIFICATION METHODS: S
 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

55 AT AAT GCA CCC CGG ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC 47
 Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser
 40 45 50 55

CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT 95
 Gln Pro Arg Gly Met Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile
 60 65 70
 5 TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT TCC TTT AAG 134
 Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile Ser Phe Lys
 80 85

(6) INFORMATION FOR SEQ ID NO:5:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 87 base pairs
 (B) TYPE:nucleic acid
 (C) STRANDEDNESS:double
 (D) TOPOLOGY:linear
 15 (ii) MOLECULE TYPE:Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM:human
 (F) TISSUE TYPE:placenta
 (ix) FEATURE:
 (A) NAME/KEY:exon
 20 (B) LOCATION:1..87
 (C) IDENTIFICATION METHODS:S
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAATAAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG 50
 Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val
 -35 -30 -25
 25 GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G 87
 Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
 -20 -15 -10

(7) INFORMATION FOR SEQ ID NO:6:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 base pairs
 (B) TYPE:nucleic acid
 (C) STRANDEDNESS:double
 (D) TOPOLOGY:linear
 35 (ii) MOLECULE TYPE:Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM:human
 (F) TISSUE TYPE:placenta
 (ix) FEATURE:
 (A) NAME/KEY:exon
 40 (B) LOCATION:1..87
 (C) IDENTIFICATION METHODS:S
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CT GAA GAT GAT G 12
 45 Ala Glu Asp Asp Glu
 -10

(8) INFORMATION FOR SEQ ID NO:7:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2167 base pairs
 (B) TYPE:nucleic acid
 (C) STRANDEDNESS:double
 (D) TOPOLOGY:linear
 (ii) MOLECULE TYPE:Genomic DNA
 (vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(F) TISSUE TYPE: placenta

(iX) FEATURE:

(A) NAME/KEY: exon + 3' UTR

(B) LOCATION: 1..2167

(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

10	GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA	48
	Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile	
85	90 95 100	
	TTC TTT CAG AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA	96
	Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu	
	105 110 115	
15	TCT TCA TCA TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC	144
	Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp	
	120 125 130	
	CTT TTT AAA CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT	192
	Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser	
20	135 140 145	
	ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAAATTCA TGCCGGCGC	246
	Ile Met Phe Thr Val Gln Asn Glu Asp	
	150 155	
	AGTGGCTCAC GCCTGTAATC CCAGCCCTTT GGGAGGCTGA GGCAGGGCAGA TCACCCAGAGG	306
	TCAGGTGTTCAAGACCAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA	366
25	AAAATTAGCT GAGTGTAGTG ACCCATGCCCTCAATCCCAG CTACTCAAGA GGCTGAGGCA	426
	GGAGAATCAC TTGCACTCCG GAGGTGGAGG TTGTGGTGAG CCGAGATTGC ACCATTGCGC	486
	TCTAGCCTGG GCAACAAACAG CAAAACCTCA TCTCAAAAAA TAAAATAAAT AAATAAACAA	546
	ATAAAAAATT CATAATGTGA ACTGTCTGAA TTTTATGTT TAGAAAGATT ATGAGATTAT	606
	TAGTCTATAA TTGTAATGGT GAAATAAAAT AAATACCAAGT CTTGAAAAAC ATCATTAAAGA	666
	AATGAATGAA CTTTCACAAA AGCAACACAA CAGACTTTCC CTTATTTAAG TGAAATAAAT	726
30	AAAATAAAAT AAAATAATGT TTAAAAAATT CATAGTTGA AAACATTCTA CATTGTTAAT	786
	TGGCATATTA ATTATACTTA ATATAATTAT TTTTAAATCT TTTGGGTAT TAGTCCTAAT	846
	GACAAAAAGAT ATTGATATTG GAACCTTCTA ATTTTAAGA ATATCGTTAA ACCATCAATA	906
	TTTTTATAAG GAGGCCACTT CACTTGACAA ATTTCTGAAT TTCTCCAAA GTCACTATAT	966
	TTTTAAAATT CAGTTTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC	1026
35	CAACTGACAT TCATCCTAGG AATGCAAAGA TGGTTAATA TCCTAAATC AATTAAACATA	1086
	ACATACTATA TTAATAAAAGT ATCAAAACAG TATTCTCATC TTTTTTCTT TTTTCACAAAT	1146
	TCCTTGGTTA CACTATCATC TCAATAGATG CAGAAAAGC ATTTGACAAA ATCCAATCA	1206
	TAATAAAAAT TCTCAAACCTT GAAAGAGAAC ATCATAAAGG CATCTATGAA AAACCTACAG	1266
	CTAATATCAT ACTTAACGAT GAAAAACTGA ATTATTTAC CCTAAGATCA AGAATAATGC	1326
	AAGCATGTCA GCTCTTGCAA CTTCTATTCA ACATTGTACT GGAGGTTCTA GCCAGAGCAA	1386
40	CCATACAATA AATAAAAATA AAAGGCACCC AGATTAGAAA GGAAGTCTT ATTTCAGAC	1446
	AACATGGTTCA TTATGCAGA AAACCGTCAG GAATACACAC ACATGTTAGA ACTAATAAGT	1506
	TCAGCAAGGT TGCAGGTGTC AATATCAATA TGCAAAATA CATTGAAGGC TGGGCTCAGT	1566
	GGAGATGGCA TGACCTTTG GTCCCAGCTA CTTGGGAGGC TGAGGTAGGA GGATCACTG	1626
	AGGTGAGGAG TTTGAGGCTA TAGTGCATG TGATCTTGCC TGTGAATAGC CACTGCACTC	1686
	GAGCCTAGGC AACAAAGTGA GACCCCGTCT CCAAAAAAA AAATGGTATA TTGGTATTC	1746
45	TGTATATGAA CAATGAATGA TCTGAAAACA AGAAAATTCC ATTACGATG GTATTAAGAA	1806
	AATAAAATAC AAATAAATTG AGCAAAATAA TTATAAAACT TGTACATCGA AAATTTCAA	1866
	GCACCTGAG GGAAATTAAA GATGATCTAA ATAATTGGAG AGACACTCTA TGATCACTGA	1926
	TTGGAAAATT CATCAATAT TGTTAAGATA ACAATTGTCC CCAAAATTGAT GCATGCATT	1986
	AATTAGTCT TCATCAAAAT TCCAGCAGGG TTTTGAGA AATTGACAAG CTGTACCCAA	2046
	AATGTATATG GAAATGAAAA GACCCAGAAG AGCAAATAAT TTTTAAAAAA CAAAGTTGGA	2106
50	AAACTTTAC TTCCTAATT TAAAACCTA TATAAACCTA AAGTTATCAA GACCATTAG	2166
	T	2167

(9) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1334 base pairs

(B) TYPE:nucleic acid
 (C) STRANDEDNESS:double
 (D) TOPOLOGY:linear
 5 (ii) MOLECULE TYPE:Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM:human
 (F) TISSUE TYPE:placenta
 (ix) FEATURE:
 (A) NAME/KEY:inton
 10 (B) LOCATION:1..1334
 (C) IDENTIFICATION METHODS:E
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTATTTTTT	TAATTCGCAA	ACATAGAAAT	GAATAGCTAC	TTCTTCCAT	TCTGTTTAC	60
15 TGCTTACATT	GTTCCGTGCT	AGTCCAATC	CTCAGATGAA	AAGTCACAGG	AGTGACAATA	120
ATTTCACTTA	CAGGAAACTT	TATAAGGCAT	CCACGTTTT	TAGTTGGGT	AAAAAATTGG	180
ATACAATAAG	ACATTGCTAG	GGGTCATGCC	TCTCTGAGCC	TGCCTTGAA	TCACCAATCC	240
CTTTATTGTG	ATTGCATTAA	CTGTTAAAAA	CCTCTATAGT	TGGATGCTTA	ATCCCTGCTT	300
GTTACAGCTG	AAAATGCTGA	TAGTTTACCA	GGTGTGGTGG	CATCTATCTG	TAATCCTAGC	360
TAATTGGGAG	GCTCAAGCAG	GAGGATTGCT	TGAGGCCAGG	ACTTTGAGGC	TGTAGTACAC	420
20 TGTGATCGTA	CCTGTGAATA	GCCACTGCAC	TCCAGCCTGG	GTGATATACA	GACCTTGTCT	480
CTAAAATTAA	AAAAAAAAC	AAAAAAAACC	TTAGGAAAGG	AAATTGATCA	AGTCTACTGT	540
GCCTTCAAA	ACATGAATT	CAAATATCAA	AGTTAGGCTG	AGTTGAAGCA	GTGAATGTGC	600
ATTCTTAA	AATACTGAAT	ACTTACCTTA	ACATATATTT	AAATATTTT	ATTTAGCATT	660
30 TAAAAGTTAA	AAACAATCTT	TTAGAATTCA	TATCTTTAA	ATACTCAAA	AAGTTGCAGC	720
GTGTGTGTTG	TAATACACAT	TAAACTGTGG	GGTTGTTTGT	TTGTTGAGA	TGCAGTTCA	780
25 CTCTGTCACC	CAGGCTGAAG	TGCAGTGCAG	TGCAGTGGTG	TGATCTCGGC	TCACTAAC	840
CTCCACCTCC	CACGTTCAAG	CGATTCTCAT	GCCTCAGTCT	CCCGAGTAGG	TGGGATTACA	900
GGCATGCACC	ACTTACACCC	GGCTAATT	TGTATTTTA	GTAGAGCTGG	GGTTTCACCA	960
TGTTGGCCAG	GCTGGTCTCA	AACCCCTAAC	CTCAAGTGT	CTGCTGCCT	CAGCCTCCCA	1020
AACAAACAAA	CAACCCACA	GTTTAATATG	TGTTACAACA	CACATGCTGC	AACTTTATG	1080
AGTATTTAA	TGATATAGAT	TATAAAAGGT	TGTTTTAAAC	TTTTAAATGC	TGGGATTACA	1140
35 GGCATGAGCC	ACTGTGCCAG	GCCTGAAC	TGTTTTAAA	AATGTCTGAC	CAGCTGTACA	1200
TAGTCTCCTG	CAGACTGGCC	AAGTCTCAA	GTGGGAACAG	GTGTATTAAG	GACTATCCT	1260
TGGTTAAATT	TCCGCAAATG	TTCTGTGCA	AGAATTCTTC	TAACTAGAGT	TCTCATTAT	1320
TATATTATT	TCAG					1334

35 (10) INFORMATION FOR SEQ ID NO:9:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH:4773 base pairs
 (B) TYPE:nucleic acid
 (C) STRANDEDNESS:double
 (D) TOPOLOGY:linear
 40 (ii) MOLECULE TYPE:Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM:human
 (F) TISSUE TYPE:placenta
 (ix) FEATURE:
 (A) NAME/KEY:inton
 45 (B) LOCATION:1..4773
 (C) IDENTIFICATION METHODS:E
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

50 GTAAGACTGA	GCCTTACTTT	GT	TTCAATC	ATGTTAATAT	AATCAATATA	ATTAGAAATA	60
TAACATTATT	TCTAATGTTA	ATATAAGTAA	TGTAATTAGA	AAACTCAAAT	ATCCTCAGAC	120	
CAACCTTTG	TCTAGAACAG	AAATAACAAG	AAGCAGAGAA	CCATTAAAAGT	GAATACTTAC	180	
TAAAATTAT	CAAACCTTTT	ACCTATTGTG	ATAATGATGG	TTTTCTGAG	CCTGTCACAG	240	
GGGAAGAGGA	GATACAACAC	TTGTTTTATG	ACCTGCATCT	CCTGAACAAT	CAGTCTTTAT	300	
ACAAATAATA	ATGTAGAATA	CATATGTGAG	TTATACATT	AAGAATAACA	TGTGACTTTC	360	
55 CAGAATGAGT	TCTGCTATGA	AGAATGAAGC	TAATTATCCT	TCTATATTTC	TACACCTTG	420	

5	TAAATTATGA TAATATTTA ATCCCTAGTT GTTTGTTGC TGATCCTTAG CCTAAGTCTT 480 AGACACAAGC TTCAGCTTC AGTTGATGTA TGTTATTTT AATGTTAATC TAATTGAATA 540 AAAGTTATGA GATCAGCTGT AAAAGTAATG CTATAATTAT CTTCAAGCCA GGTATAAAGT 600 ATTTCTGGCC TCTACTTTT CTCTATTATT CTCCATTATT ATTCTCTATT ATTTTCTCT 660 ATTTCTCTCA TTATTGTTAG ATAAACCACA ATTAACTATA GCTACAGACT GAGCCAGTAA 720 GAGTAGCCAG GGATGCTTAC AAATTGGCAA TGCTTCAGAG GAGAATTCCA TGTCTGAAG 780 ACTCTTTTG AGTGGAGATT TGCCAATAAA TATCCGCTT CATGCCACC CAGTCCCCAC 840 TGAAAGACAG TTAGGATATG ACCTTAGTGA AGGTACCAAG GGGCAACTTG GTAGGGAGAA 900 AAAAGCCACT CTAAAATATA ATCCAAGTAA GAACAGTGCA TATGCAACAG ATACAGCCCC 950 10 CAGACAAATC CCTCAGCTAT CTCCCTCCAA CCAGAGTGCC ACCCCTTCAG GTGACAATT 1020 GGAGTCCCCA TTCTAGACCT GACAGGCAGC TTAGTTATCA AAATAGCATA AGAGGCCTGG 1080 GATGGAAGGG TAGGGTGGAA AGGGTTAAGC ATGCTGTTAC TGAAACAACAT AATTAGAAGG 1140 GAAGGAGATG GCCAAGCTCA AGCTATGTGG GATAGAGGAA AACTCAGCTG CAGAGGCAGA 1200 TTCAGAAAATC GGGATAAGTC CGAACCTACA GGTGGATTCT TGTTGAGGG GACTGGTGAA 1260 AATGTTAAGA AGATGGAAAT AATGCTTGGC ACTTAGTAGG AACTGGCAA ATCCATATT 1320 15 GGGGGAGCCT GAAGGTTATT CAATTGAT GGCCCTTTA AATAAAAAGA ATGCTGGCTGG 1380 GCGTGGTGGC TCACACCTGT AATCCCAGCA CTTGGGGAGG CCGAGGGGG CGGATCACCT 1440 GAAGTCAGGA GTTCAAGACC AGCCTGACCA ACATGGAGAA ACCCCATCTC TACTAAAAT 1500 ACAAAATTAG CTGGGCGTGG TGGCATATGC CTGTAATCCC AGCTACTCG GAGGCTGAGG 1560 CAGGAGAACAT TTTGAAACC GGGAGGCAGA GTTGTGCGATG AGCCTAGATC GTGCCATTGC 1620 20 ACTCCAGCCT GGGCAACAAG AGCAAAACTC GGTCTCAAAA AAAAAAAA AAAAGTGAAG 1680 TTAACCAAAG GCATTAGCTT AATAATTAA TACTGTTTT AAGTAGGGCG GGGGGTGGCT 1740 GGAAGAGATC TGTGTAATG AGGGAATCTG ACATTTAACG TTCATCAGCA TCATAGCAAA 1800 TCTGCTCTG GAAGGAACTC AATAAATATT AGTTGGAGGG GGGGAGAGAG TGAGGGTGG 1860 ACTAGGACCA GTTTTAGCCC TTGTCCTTAA TCCCTTTTCC TGCCACTAAT AAGGATCTTA 1920 GCAGTGGTT TAAAAGTGGC CTAGGTTCTA GATAATAAGA TACAACAGGC CAGGCACAGT 1980 25 GGCTCATGCC TATAATCCC GCACTTTCGG AGGGCAAGGC GAGTGTCTCA CTTGAGATCA 2040 GGAGTTCAG ACCAGCCTGG CCAGCATGGC GATACTCTGT CTCTACTAAA AAAAAATACAA 2100 AAATTAGCCA GGCATGGTGG CATGCACCTG TAATCCCAGC TACTCGTGAG CCTGAGGCAG 2160 AAGAATCGCT TGAACACCAGG AGGTGTAGGC TGAGATCGCA CCACTGCAC 2220 CCAGCCTGGG CGACAGAAATG AGACTTGTG TCAAAAAAAAG AAAAGATAC AACAGGCTAC 2280 CCTTATGTGC TCACCTTCA CTGTTGATTA CTAGCTATAA AGTCTATAA AGTTCTTGG 2340 30 TCAAGAACCT TGACAACACT AAGAGGGATT TGCTTGAGA GTTACTGTC AGAGTCTGTT 2400 TCATATATAT ACATATACAT GTATATATGT ATCTATATCC AGGCTTGGCC AGGGTTCCCT 2460 CAGACTTCC AGTGCACTTG GGAGATGTTA GGTCAATATC AACTTCCCT GGATTAGAT 2520 TCAACCCCTT CTGATGTAAA AAAAAAAA AAAAGAAAG AAATCCCTT CCCCTTGAG 2580 CACTCAAGTT TCACCCAGGT GGGCTTCCA AGTTGGGGT TCTCCAAGGT CATTGGGATT 2640 GCTTTCACAT CCATTGCTA TGACCTTCC CTATGATGGC TGGGAGTGGT CAACATCAAA 2700 35 ACTAGGAAAG CTACTGCCA AGGATGTCCT TACCTCTATT CTGAAATGTC CAATAAGTGT 2760 GATTAAGAAG ATTGCTGTG CTACCTATCC ACACTCTCGC TTTCACCTGT AACTTTCTTT 2820 TTTCTTTT TTCTTTTTT CTTTTTTT GAAACGGAGT CTCGCTCTGT CGCCCAAGGCT 2880 AGAGTGCAGT GGCACGATCT CAGCTACTG CAAGCTCTGC CTCCCGGGTT CACGCCATT 2940 TCCTGCTCA CCCTCCCAAG CAGCTGGGC TACAGGGCCC TGCCACCATG CCCAGCTAAT 3000 TTTTGTATT TTTAGTAGAG ACGGGGTTTC ACCGTGTTAG CCAGGATGGT CTCGATCTCC 3060 40 TGAACCTGTG ATCCGCCCGC CTCAGCCTCC CAAAGTGTG TGATTACAGG CGTGAGGCAT 3120 CGCACCCGGC TCAACTGTAA CTTCTATAC TGTTCTATCT TCCCTGTAA TGTTACTAGA 3180 GCTTTGAAG TTTGGCTAT GGATTATTC TCATTTATAC ATTAGATTTC AGATTAGTTC 3240 CAAATTGATG CCCACAGCTT AGGGTCTCTT CCTAAATTGT ATATTGTAGA CAGCTGCAGA 3300 AGTGGGTGCC AATAGGGGAA CTAGTTATA CTTTCATCAA CTTAGGACCC ACACTTGTG 3360 ATAAAAGAAC AAGGTCAAGA GTTATGACTA CTGATTCCAC AACTGATTGA GAAGTGGAG 3420 45 ATAACCCCGT GACCTCTGCC ATCCAGAGTC TTTCAGGCAT CTTGAAGGA TGAAGAAATG 3480 CTATTTAAAT TTGGAGGT TCTCTATCG TGCTTAGGAT CATGGGAATC TGTGCTGCCA 3540 TGAGGCCAAA ATTAAGTCCA AAACATCTAC TGTTCCAGG ATTAACATGG AAGAACCTTA 3600 GGTGGTGCCC ACATGTTCTG ATCCCATCCTG CAAAATAGAC ATGCTGCACT AACAGGAAAA 3660 GTGCAGGCAG CACTACCAGT TGGATAACCT GCAAGATTAT AGTTCAAGT AATCTAACCA 3720 50 TTTCTCACAA GGCCTATTTC TGACTGAA ACATACAAGA ATCTGCATT TGCCTCTAA 3780 GGCAGGGCCC AGCCAAGGGAG ACCATATTCA GGACAGAAAT TCAAGACTAC TATGGAACGT 3840 GAGTGCTTGG CAGGGAAAGAC AGAGTCAAGG ACTGCCACT GAGCCAATAC AGCAGGCTTA 3900 CACAGGAACC CAGGGCCTAG CCCTACAACA ATTATTGGGT CTATTCACTG TAAGTTTAA 3960 TTTCAGGCTC CACTGAAAGA GTAAGCTAAG ATTCCCTGGCA CTTCTGTCT CTCTCACAGT 4020 TGGCTCAGAA ATGAGAACTG GTCAGGCCAG GCATGGTGGC TTACACCTGG AATCCCAGCA 4080
---	--

5 CTTTGGGAGG CCGAAGTGGG AGGGTCACTT GAGGCCAGGA GTTCAGGACC AGCTTAGGCA 4140
 ACAAAAGTGAG ATACCCCCCTG ACCCTTCTC TACAAAATA AATTAAAGA ATTAGCAGGAA 4200
 TGTGGTGGTG TATACTTACA GTCCAGCTA CTCAGGAGGC TGAGGCAGGG GGATTGCTTG 4260
 AGCCCAGGAA TTCAAGGCTG CAGTGGCTA TGATTTCACTC ACTGCACCTTC TGGCTGGCA 4320
 ACAGAGCGAG ACCCTGTCTC AAAGCAAAA GAAAAGAAA CTAGAAGTAG CCTAAGTTG 4380
 TGGGAGGAGG TCATCATCGT CTTTAGCCGT GAATGGTTAT TATAGAGGAC AGAAATTGAC 4440
 ATTAGCCAA AAAGCTTGTG GTCTTGTG GAACTCTACT TAATCTTGAG CAAATGTGGA 4500
 CACCACTCAA TGGGAGGAGA GAGAAGTAAG CTGTTTGATG TATAGGGGAA AACTAGAGGC 4560
 CTGGAACATGA ATATGCATCC CATGACAGGG AGAATAGGAG ATTGGAGTT AAGAAGGAGA 4620
 10 GGAGGTCACT AGTGTGTTC AGAGATTTT TTTATGTAAC TCTTGAGAAG CAAAATCT 4680
 TTTGTTCTGT TTGGTAATAT ACTTCAAAAC AAACCTCATA TATTCAAATT GTTCATGTCC 4740
 TGAAATAATT AGGTAATGTT TTTTCTCTA TAG 4773

(11) INFORMATION FOR SEQ ID NO:10:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8835 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 20 (ii) MOLECULE TYPE: Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: human
 (F) TISSUE TYPE: placenta
 (ix) FEATURE:
 (A) NAME/KEY: intron
 (B) LOCATION: 1..8835
 25 (C) IDENTIFICATION METHODS: E
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

30 GTAAGAAATA TCATTCCTCT TTATTTGAA AGTCAGCCAT GGCAATTAGA GGTAAATAAG 60
 CTAGAAAGCA ATTGAGAGGA ATATAAACCA TCTAGCATCA CTACGATGAG CAGTCAGTAT 120
 CAACATAAGA ATATATAAGCA AAGTCAGAGT AGAATTTTT TCTTTTATCA GATATGGGAG 180
 AGTATCACTT TAGAGGAGAG GTTCTCAAAC TTTTGCTCT CATGTTCCCT TTACACTAAG 240
 CACATCACAT GTTAGCATAA GTAACATTTT TAATTAAAAA TAACTATGTA CTTTTTTAAC 300
 AACAAAAAAA AGCATAAAGA GTGACACTTT TTTATTTTA CAAGTGTGTTT AACTGGTTA 360
 ATAGAAGCCA TATAGATCTG CTGGATTCTC ATCTGCTTTG CATTTCAGACT ACTGCAATAT 420
 35 TGACACAGAAT GCAGCCTCTG GTAAAACCTTG TTGTACACTC ATGAGAGAAT GGGTGAAAAA 480
 GACAAATTAC GTCTTGAAT TATTAGAAAT AGCTTTCACT TTAGGAACCT CCTGAGAATT 540
 GCTGCTTTAG AGTGGTAAGA TAAATAAGCT TCTCTTTAAA CGGAATCTCA AGACAGAAC 600
 AGTTACATTA AAAGCAAACA AAAAATTGTC CCATGGTTAG TCATCTGTG AAATCTGCCA 660
 CACCTTGGG CTGGGCTACA ATTGGATAAT ATAGCATTCC CCGAGATAAT TTTCTCTCAC 720
 AATTAAGGAA AGGGCTGAAT AAATATCTCT GTTTGAAGTT GAATAACAAA AATTAGGACC 780
 40 CCCTAAATT TAGGGCTCCT GAAATTGTC TTTTGCTTA TATTCACTA CTTTACGTT 840
 TATTAATCT TCTTTCAGGC CAGGTGCACT AGCTCATGCC TAGAATCTCA GGCAGGCCTG 900
 AGCCCAGGAA TTGAGACCA GCCAGGGCAA CACAGTCTC ACAAAATAAT AAAAAATTAC 960
 CTGGGTGTGT TGGTGCATGC CTGTAGAACT ACTCAGGATG CTGAGGACTG CTTGAGCCA 1020
 GGATAGCCAA ATCTGTGGTG AGTCAGCCA CTAAACAGAG CGAGACTTT TCACAAAC 1080
 AAACAAAAAA ACAAAACAAAC TTCCCTCAAA ATAACTTTT ATCTGCAATG TTTTCCTATT 1140
 45 GCCTGTGAGA TAAATTTAC TCTTTACCT GATTTCAAA GCCCTCCATA ATCTAATCCG 1200
 ACTTTACCTT GTGTTCACTG CAAAATAGCA GGACTGTTCC ACTACAATCC AAAAATCACA 1260
 GGTTGGGTGC AGTGGCTCAC TCCTGTAATC CCAACACTT GGAAGGCCAA GGCAGGTGGA 1320
 TTGCTTCAGC TCAGGAGTTC AAGACCAGCC TGGGCAACAT GGCAAAACC CTGTCTCTCC 1380
 AAAACATACA AAAATTAGCC AGATGTGGTA GTATGTGCT GTAGTCCAA CTACTCAAA 1440
 50 GGCTAAGGCA AGAGGATCAC TTGAGGCCAG GAGGTCAAGG CTACAGTGAG CCATGTTAC 1500
 TGTGTCAGT CACTCCAGCC TGGGTGATAG AGCAAGACCA TGTCTCAAA AAAAAAAA 1560
 GAAAAGAAA GAAAAAAACA TCGCTTATT CAGTTCACCC CCACCCACAC ATTGTTTGA 1620
 TTATCACATA AATGCTGGTC CATTGCTTC TCTATCTATT CAAATTTTA AGCATTCTT 1680
 GAGATTCAAC TCAATTCTCC TTTTCAAAC AGGCCATTAA AACTACATCA GTTCCATT 1740
 GATTTCTTG CTTTGAGTCT ACAGACTCAA AAACAAAAAC TTAAAAACTT ATTTTTAAG 1800
 TTTCTGCTA CTCTCACTTC TTCAACACTC ACATACACGC ATTCAATAAGA AGATGGCAGA 1860

	ATGTTCAAGG	ATAAAATGAT	TTATAGAACT	GAAAAGTTAG	GTTTGATCT	TGTTGCTGTC	1920
5	AAGATGACTA	CCTACCTGAT	CTCAGGTAAT	TAATTATGTA	GCATGCTCCC	TCATTTCATC	1980
	CCATAACCTAT	TCAACAGGAT	TGGAATTCCA	CAGCAAGGAT	AAACATAATC	ATAGTTGCTT	2040
	TTCAAGTTCA	AGGCATTTA	ACTTTTAATC	TAGTAGTATG	TTTGTGTTG	TTGTTGTTGT	2100
	TTGAGATGGA	GCCCTGCTGT	GTCACCCAGG	CTGGAGTGCA	GTGGCACGAA	CTCGGCTCAC	2160
	TGCAACCTCT	GCCTCATGGG	TTCAATCAGT	TATTCTGCCT	CAGTGTCCCA	AGTAGCTGGG	2220
	ACTACAAGGC	ACATGCCACC	ATGCTGGCT	AATTTTGTA	TTTTTAGTAG	AAACAGGGCT	2280
10	TCACCATGTT	GGCCAGGCTG	GTCTCGAACT	CCTGACCTCA	AGTGTACAG	CCGCCTCGGC	2340
	CTCCCCAAAGT	GCTGGGATTA	CAGGCATAAG	CCACCGTGC	CAGGCTAATA	GTATGTTTT	2400
	AAACTCTTAG	TGGCTTAACA	ATGCTGGTG	TATAATAAAT	ATGCCATAAA	TATTTACTGT	2460
	CTTAGAATT	TGAAGAAGTG	GTTACTAGGC	CGTTGCCCC	ATATCAATGG	TTCTCTCCTT	2520
	ACAGCTTAA	TTAGAGTCTA	GAATTGCAGG	TTGGTAGAGC	TGGAACAGAC	CTTAAAGATT	2580
	GACTAGCCAA	CTTCCTTGTG	CAAATGAGGG	AACTGAGACC	CTTAAATTAA	AGTGACTTGC	2640
15	CCCAGACAAA	ACTGGAAC	ATGTCCTA	ATTCACATCA	TGAAATTCTA	CCATTCACTA	2700
	GCCTCTGGCT	AGTTGTCAA	GTATTGCATA	ACTAAATT	TATGCTGTT	TTAAAGAAC	2760
	AATTGTCACT	GTCTACTCCT	GGGAGGGTCT	TTCTGAGGTG	GTTTATAACT	CTTAAAAAAA	2820
	AAAAAGTCAG	TAGTCTGAGA	ATTTAGACG	AAATAGTCAA	AGCATT	TCCAATGGAT	2880
	CTATAATT	CATAGATTAG	AGTTAAATCA	AAGAACACG	GATGAGAAAG	GAAGAGGAAA	2940
	ATTGAGGAGA	GGAGGAATGG	GGATGAGAAC	ACACTACTG	TAATCAGTCA	TAGATGTACT	3000
20	GAGAACAAC	AAGAAGAATT	GTAAGAAAAT	AAGAATGAAG	AATTCAAAAT	CAACACATGA	3060
	AATAAAAAGA	AACTACTAGG	GAAAATGGA	GAAGACATTA	GAAAATTAT	TCTATT	3120
	AAATTCTGTT	TTCAGGCTTC	CCTCCGTGTC	TTCCCTCCTC	TCATTGGTT	TCAGGTGGAG	3180
	GGAAAGTTA	AGATGGAAA	AATATATA	TTCTACACAT	CCCTTCTAC	GCTGTTGTCA	3240
	TGGCAACAAG	GTTTATCATA	GCAAATT	ATTCAACAA	CATTATTGA	GTTCTTACTG	3300
	TGTGGTAAGC	TCTTCCAGG	TGTTAAAAT	TCAGGGAAA	AAAGACAAC	CATTGTCTTA	3360
25	AAACTCAGAT	GAAAGCTGAA	CAGACCTATT	TTAATCAA	GTAATCTAA	TTAGGGTAG	3420
	TAAGAGCTAT	TTAAGAAC	TGAACAGGTG	TGAAGGAGGT	AGGACTCTGA	GGAGAGAATA	3480
	GTTAGCTAGG	AATGAAAGAG	CAGAGAAGTT	TTCTAGAGG	AACTATTAA	GCTGGAGTT	3540
	ACGGGATGAA	AGATGAGGC	GGGTTGCGAG	GCAAAAAAAA	AAAAAAGCA	GGGGAAGGGG	3600
	AAGTTCTGGC	CTGGCAGAGA	GAATAACTGT	GGCAACAA	GAGGAGAGTC	TGGAAGCAAG	3660
	AAAACCAAGT	AGAAGAGTAT	AAAATAGAA	GATGCCAGGG	GTAATGAGGG	CTTGATTAA	3720
30	AACAGTGCTG	TGAGGAGATGG	AGAGGAGATA	CCAAATTCTG	GAGACATT	TGAGTTAGAA	3780
	CCTACAGT	TTATCAGACA	AGGGAAAGAT	TAGACAAAGG	AGTTAAGAAT	GACTCCCAGG	3840
	TTTCAGTTG	GGGCAGGTAA	CTAGGACATG	TTTGAAAAG	TAATGTTATTG	GATCTCTTAC	3900
	CATTGAAACT	ATGTATGTG	AGCCAAATT	AAATTGTAC	ATGTATATAA	CTCTCCCCC	3960
	ACCACCACTA	ACTACTTCCC	TAACTCTCA	CTTGATGCC	AGACTTCTA	AAAGAATAGT	4020
	TTGTAGTCAC	TGTCTTACT	TTTCCCTC	CATTCTGTCC	TAGATATTG	TCCACCTACC	4080
35	ATCTGCTGCC	TCCACTTTAC	CCAAACTGTT	CTACGGTTG	CCAAAACCT	CTAATTGCCA	4140
	AATTCAATGA	ACAAGTTAA	GCTTATATGT	AAATTAGGAG	CTCTACAGTT	TGATTCGAG	4200
	CAGCCCCCTCC	TGAAACCCCTT	TCTCTTCGA	CTTCTGTGAC	ACATCTCAGA	TTTACAAAC	4260
	TGAACAAATT	ATTTTACACT	TGAGCTGTAT	TTTCGTTCTT	CTTTCTTGAT	GAATGAGGTA	4320
	ACCACTCAAC	AAATTGCCCA	AGCCAAAAC	TACGAAGTCA	TCCTCAGTTC	CTCCTTCTTC	4380
	TGTTTGACCC	ACAACAGATC	AGCTGAGAAA	TCCCGCTGTT	TAGTATCTCT	TGAATTCTATT	4440
40	ACCTTAATT	ATAGCCTCAT	CAACTCTAA	TTGTTAAAAT	TACTTCAGTA	GTTGTTGTCT	4500
	GACCTCTGTC	CAATCTGTT	CAATCAGGTC	CATTCTT	TTCTGGTGG	TGGTGGTGGT	4560
	GTTGACAGAG	TTTCGTTTT	GCTGCCAGG	CTGAAGTCA	GTGGAGCACT	TCACTGCAAC	4620
	CACAGCCTCC	TGGGTTTAAG	CAGTTCACCC	TCCCGAGTAG	CTGGGACTAC	AGGTATGTGC	4680
	CACCAACCCC	AGCTAATT	GTGTTTCA	TAGAGACAGG	GTGGTACCAT	GTTGGTCAGG	4740
	CTGGTCTCAA	ACTCCTGACC	TCAAGAAC	CACCCACCTC	AGCCTCCCAA	AGTGTGGGA	4800
45	TTACAGGCAT	GAGCCACTGC	ACACGGACCA	GATCCATTG	TTATGTTGCT	TCTAGAGTGA	4860
	GTTTTAAAAA	CACAAATTG	ACCATATCTT	TCTCCAATT	AAGTCAGTAT	TTTTTTTTTC	4920
	AGGAAAAAAAC	AGTTCAAAC	CTTCTAGTCTG	CTTACACAA	GCCTTGTTAG	TCTGACTCTT	4980
	CTTTCCAAGC	TTTCATCAA	GTATACTGCA	AGTTACATT	TATGTGAATT	GAATTAGGCA	5040
	ACGGTATAAA	AATTATAGTT	TATATGGCA	AAATGGAAAT	AATGTTAATC	CTTCCAATA	5100
50	GTTTATCTAG	AATGACATAA	TTTCAAAGCT	GTCAGGTCAA	ATGAGTTATA	AACTGTTAAC	5160
	ACTATTGCCA	CATGCAAGTG	TCTCTTATAC	TTGGTAAAT	TATCTGCTTC	CATGTCATTA	5220
	TTATGAAAT	TAGACTTTAA	ATAACTCAGA	AGTTCTTCAG	ACATACAGGT	TATTATTGTTG	5280
	CTTTTTAAAC	ATAATT	ATAATT	ATATGATAAT	GTTATCCAAG	TGCTAAGGGA	5340
	TGTATTGTTA	CTGCTGTGCA	AAAAAAA	AAAAAAAAC	TCCAAATAAA	TATGTTGAA	5400
	CCAAGTTAT	ATGCAAGAAA	ACAATTAA	AAAGGCCAAA	GTACCAAC	AATAGGCTGT	5460
	GTGGAGACGG	CAGGCTACAA	AAACACTAGTA	ATAATGCTGA	GAAAGTTGAA	AAAAGAAAGA	5520

	AAGCAACAAT	ATGCTTGTT	TGTTGTTAGGT	TTATGTTACTC	CAAGAATATC	TCCTCTCAAA	5580
5	CTTTTACGTT	TTTTCCAAG	AAAAGTTAAC	TTTGGCTGGG	CGCAGTGGCT	CTTGCCCTGTA	5640
	GTCCCAGCCT	TTGGGAGGCC	AAGGCCGGCA	GATCACCTGA	GGTCAGGAGT	TTGAGACCAG	5700
	CCTGACCAAA	AATGGAGAAA	CCCGCCCCCCC	TCACTACTAA	AAGAATACAA	AATTAGGCCG	5760
	GGCACAGTGG	CTTACCCCTG	TGATCCCAGC	ACTTTGGGAG	GCCGAAGCAG	GAAGATCACC	5820
	TGAGGTCAAGG	AGTTCGAGAC	CAGCCATGGA	GAAACCCGTC	TCTACTAAAAA	ATACAAAATT	5880
	AGCCGGCGT	GGTGGTGCAT	GACTGTAATC	CCAGCTACTC	AGGAGGCTAA	GGCAGAGAAT	5940
	CACTTGAACC	CAGGCAGTGG	AGGTTGCAGT	GAGCCGAGAT	CGTGCCTTG	CACTCCAGCC	6000
	TGGGCAACAA	GAGCGAAACT	CTGTATCCAA	AAAACAAAAG	AAAAGAAAAG	GTAACCTTGA	6060
10	ACTATGTGAG	ATCTTTAGAA	ATGCATTCTT	TCTGTAAAAT	GTGACTACAT	TTGCCTTATT	6120
	TATGGTAAAAA	ATGTTGAGGC	CTCAAACAAAC	CCATATTTC	TCGGTCTCCC	CGCTGCCTAG	6180
	CCTTTGTTCA	CATTGCTCT	TCTTGGTGGG	AGCTCTTCT	CTGGCCTTGA	AAATGCCTGC	6240
	TTCTCTTCA	AGGTAGCACA	GTCATCACTT	TCTGTGGTAA	CCTTCTCCAG	CACCATCAA	6300
	CAGAAAGAAT	GAATCTCTG	TAATTCAGC	TCTTACGTCA	TTCATTACAT	TATTTGTAA	6360
15	CTCTTATAG	ATTCTTCTCT	CCCACTAGAC	TCTGAGTCAC	TGGAGAGTAG	GAGCCAACTC	6420
	TCATTCTATG	GTGGGTTGGT	CAGCTACTGG	CCACATTCT	GATGCTAGT	TAATGCTCAA	6480
	ACCTTAACGT	GTGAATCAGC	TCAAATATTG	TCCTTCTCTA	AATCCATTCA	CTCATTGACT	6540
	AACTATGTAC	TCAAACATAGT	AAACACCGT	AATTTAATCC	AATTCCTGCC	CATACTGCTT	6600
	GGTACATTTC	AGGTGAATTA	GTGGATAAA	TATGTGTAA	TTACATAATA	TTAAAGTATG	6660
	TACAGAAGAT	CATGCTAAATC	ATAATTCAAA	ACTGATAACT	AATCAAACAT	AAATGCTCTC	6720
20	AGGTTAACAA	ATGTCTGCCT	TCTCAGTTAA	TGCAGTCATT	AACAAACACC	TTCTGATGCT	6780
	GATAATAGGG	CCTTGTTCAG	CAATGAAGCC	ATAAAGGTGA	ATAAAGAACAA	TGCCCTCGTG	6840
	GAGCTCACAG	CCTAGTCATT	ATTGTTCTGA	TTTTTAATAT	TAATGTTGGT	TTGGGTTTTG	6900
	GTGAAAATG	TTTAGACTTA	TCTTAGTGAT	CTTTTCATCC	TTTGCTATAT	TATTTTCTC	6960
	TAAGAGTCTT	CCTTATCCCC	TCCTTTAAAAA	AACTAGGTGA	TAATTCTAAA	TTGTAATTT	7020
	AAATATTATA	AATAGCTTAT	AAAATTAAAT	ATTTATAATA	TTTAAATGTT	TGATAAAATAT	7080
25	TTAAATTATA	TAATATTAA	ATGTTTATTT	AAATTCAATT	GTACATCAGT	TTTTATTTTA	7140
	TTTAAATGTC	TTGGCCAGGC	ATGGTGGCTG	ACACCTTAA	TCCCAGAACT	TTGAGAGGCC	7200
	AAGTCAGGC	AACCATTG	GCTCAGGAGT	TTGAGACCA	CCTGGGCAAC	GTGGTAAAC	7260
	CCTGCTCTA	CCAAACATAT	AAAAACATTAT	TCGGGTGTGG	TGGCACGCAT	CTGTGGTCCC	7320
	AGATGGGAGT	CCCAGGCTAA	GATGGGAGAA	TCGCTTGAAC	CCAGGTGAGA	GGGGTGGGGT	7380
	GGATGTTGCA	GTGAGCTGAG	ATCGTGCAC	TGCACTCCAA	CCTGGGTGAC	AGAGTGAGAC	7440
30	TCCATCTCAA	AAAAAA	TGTTATCTAA	ATAAGATAAA	TTTAATAACT	TTTCGCACTT	7500
	AGATGAGCAT	AGGGAACTAA	ACCTAGATAA	AACTATCAA	TAAGGCCTGG	GTACAGTGAC	7560
	TCATGCCTGT	AATCTCAAGC	ACTTTGGGAG	GCCAAAATTA	TACAAAGTTA	GTTGTATAAC	7620
	ACCAACTAAC	AACTATTTG	GGGTTAGCTT	AATTCAAGATT	AATTTTTTT	AAACTGAGTT	7680
	TTAAATTCTC	GCTTACTCTA	CCATACATGC	TAGGCCTCAT	ATTATGCTAG	AAAAATTG	7740
	AGCACAGATT	TATGAATACT	CTCCTGCATA	CCATTAAATT	TTTAAACAAA	TTTAATGCA	7800
35	GTATATATGT	GCCTTTTTAC	CAACACATTA	ATAATAAGA	TCTACTGTGA	GGACTAAATT	7860
	TCTGTAATT	CAAAGTAGTA	ATGAGTTAA	ACCATGTC	AAGATCTCTG	CAATAACTGT	7920
	AGCACAACAG	AAAATAGGTA	TTCTTATTAA	TGACAGAGTC	ACAAGTACTA	CTAATAATAC	7980
	TGTGGTTTGT	TTCCCTGCAAC	TAATCATGGG	AGGAATGCTA	AATTTCAGAG	GTTGGTAAA	8040
	ATACATGTGT	ATTTTTTCC	CCATCCAAGT	TCACAGATT	CTCACACTGA	GAACTCCTAT	8100
	TCCATAACAA	AATTCTGGAA	GCCTGCACAC	CGTATTGGAA	GAAGGGCAGA	AAGGAAAAGC	8160
40	AAATGGAAGG	ATTTAAATT	TTTCAAATC	CTGTATCCCT	TGATTTACA	GCAAGATGT	8220
	ATTTATGTAT	TACTTGTGTT	AAAAATATAG	TATAATCGAG	ACTCAGATC	AAAAATCACC	8280
	GCAGCTCAGG	GAGAAAGAGG	GCCACCAAAT	GCCAGAGCCC	TTCAGCCTTC	TCCCACCTG	8340
	CCTGTACCCCT	CAGATGGAAG	CACTTTTTA	TCATTGTTTC	ACCTTAGCA	TTTGACAAT	8400
	GAAGTCACAA	ACCTTCAGCC	TCTCACCCAT	AGGAACCCAC	TGGTTGTAAG	AGAAGGATGA	8460
	AGCCAGTCCT	TCCTAAAGGG	CACGATTAGA	TGTGTTTATG	GCATCCTCAG	GTGAAACTAT	8520
45	ATTTATATTG	ACAATATATT	TATATTCTC	AAGGAATACT	AGAATAATGA	TTCAGTTCAG	8580
	TACTAGGCCA	TTTATCTACC	CTTTATAATA	TTGTTTAATG	AGAAAATGCT	TTCTATCTTC	8640
	CAAATATCTG	ATGATTTGTA	AGAGAACACT	TAAACATGGG	TATTCTATAAG	CTGAAACTTC	8700
	TGGCATTTAT	TGAATGTCAA	GATTGTTCAT	CACTATACTA	GGTGAATTAAC	TGACCACTG	8760
	ACTTGAAGGT	AGTATAAAAGT	AGTAGTAAAAA	GGTACAATCA	TTGTCTCTTA	ACAGATGGCT	8820
	CTTTGCTTTC	ATTAG					8835

(12) INFORMATION FOR SEQ ID NO:11:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1371 base pairs

(B) TYPE:nucleic acid
 (C) STRANDEDNESS:double
 (D) TOPOLOGY:linear
 5 (ii) MOLECULE TYPE:Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM:human
 (F) TISSUE TYPE:placenta
 (ix) FEATURE:
 (A) NAME/KEY:intro
 10 (B) LOCATION:1..1371
 (C) IDENTIFICATION METHODS:E
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTAAGGCTAA	TGCCATAGAA	CAAATACCA	GTTCAGATAA	ATCTATTCAA	TTAGAAAAGA	60
TGTTGTGAGG	TGAACATATTA	AGTCACTCTT	TGTGTCACCA	AATTCACTG	TAATATTAAT	120
15 GGCTCTTAAA	AAAATAGTGG	ACCTCTAGAA	ATTAACCCACA	ACATGTCCAA	GGTCTCAGCA	180
CCTTGTCA	ACA	CCACGTGTCC	TGGCACTTTA	ATCAGCAGTA	GCTCACTCTC	240
TAAGTGAC	ACA	TCATGAAAAT	CCCAGTTTC	ATGGGAAAAT	CCCAGTTTC	300
CATGGGAAA	AT	ATCCCAGTAC	AAAACGGGT	GCATTCA	AATACAATT	360
ATTGGCAA	AT	TATGTAAGAG	ATTCTCTAAA	TTTAGAGTTC	CGTGAATTAC	420
20 GTAAATATGT	TTGACAAGTA	AAAATTGATT	CTTTTTTTTT	TTTCTGTTG	ACCATT	480
AGTCAGTGG	CACAATCTCT	GCTCACTGCA	ACCTCCACCT	CCTGGGTTCA	AGCAATTCTC	540
CTGCCTCAGC	CTTCTGAGTA	GCTGGGACTA	CAGGTGCATC	CCGCCATGCC	TGGCTAATT	600
TTGGGTATTT	TTACTAGAGA	CAGGGTTTTG	GCATGTTGTC	CAGGCTGGTC	TTGGACTCCT	660
GATCTCAGAT	GATCCTCCTG	GCTCGGGCTC	CCAAAGTGCT	GGGATTACAG	GCATGAACCA	720
CCACACATGG	CCTAAAATT	GATTCTTATG	ATTAATCTCC	TGTGAACAAT	TTGGCTTCAT	780
25 TTGAAAGTTT	GCCTTCATT	GAACCTTCA	TTTAAAAGCC	TGAGCAACAA	AGTGAGACCC	840
CATCTCTACA	AAA	AAACTGCA	AAATATCCTG	TGGACACCTC	CTACCTTCTG	900
AGCAGGAGGA	TCACTTGAGC	CTAGGAATT	GAGCCTGCAG	TGAGCTATGA	TCCCACCCCT	960
ACACTCCAGC	CTGCATGACA	GTAGACCCCTG	ACACACACAC	ACAAAAAA	ACCTTCATAA	1020
AAAATTATTA	GTGACTTTT	CTTAGGTGAC	TTTCCGTTA	AGCAATAAT	TTAAAAGTAA	1080
AATCTCTAA	TTTAGAAAAT	TTTCTGTTAG	TTACATATG	AAATTTTAA	ACCTCTAGTT	1140
30 TAAGTTTTAT	GTCTAAATT	CCTGAGAAC	CACTAAGTCT	GATAAGCTC	ATTTTATGGG	1200
CCTTTGGAT	GATTATATAA	TATTCTGATG	AAAGCCAAGA	CAGACCCCTA	AACCATAAAA	1260
ATAGGAGTTC	GAGAAAGAGG	AGTAGCAAAA	GTAAAAGCTA	GAATGAGATT	GAATTCTGAG	1320
TCGAAATACA	AAATT	TATTCTGTTT	CTCTCTTTT	CCCCCTCTTA	G	1371

35 (13) INFORMATION FOR SEQ ID NO:12:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3383 base pairs
 (B) TYPE:nucleic acid
 (C) STRANDEDNESS:double
 (D) TOPOLOGY:linear
 40 (ii) MOLECULE TYPE:Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM:human
 (F) TISSUE TYPE:placenta
 (ix) FEATURE:
 (A) NAME/KEY:intro
 45 (B) LOCATION:1..3383
 (C) IDENTIFICATION METHODS:E
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GTAAAGTAGA	AATGAATT	TTTTCTTTG	CAAACTAAGT	ATCTGCTTGA	GACACATCTA	60
TCTCAC	CATT	GTCA	GTGAG	GGAAAAAAA	AATGGTTCTC	120
50 AAGAAATGTG	GA	CTCAGTAG	CACAGCTTG	GAATGAAGAT	GATCAT	180
AGAACCTCTA	GC	AAAAGATG	CTTCTCTATG	CCTTAA	AGA	240
ACAAAATAGA	CTT	GGCTGT	TTCA	TTAGATTAG	TTCTCCAGCT	300
TGTAGGGGGA	GG	GTG	AG	CATGAAGCCA	CTTGAAT	360
AACACCTCCT	CTC	AGAAATG	TTT	GGAAAG	GGATTCTGT	420

5	GTGGGGCAGA AAATTCTGGA AGTAGAGGAG ATAGGAATGG GTGGGGCAAG AAGACCACAT 480
	TCAGAGGCCA AAAGCTGAAA GAAACCATGG CATTATGAT GAATTCAAGG TAATTCAAGAA 540
	TGGAAGTAGA GTAGGAGTAG GAGACTGGTG AGAGGAGCTA GAGTGTAAA CAGGGGTGAG 600
	AGCAAGACGT TCTCTCACCC CAAGATGTGA AATTGGACT TTATCTTGGAA GATAATAGGG 660
	TTAATTAAAGC ACAATATGTA TTAGCTAGGG TAAAGATTAG TTTGTTGAA CAAAGACATC 720
	CAAAGATACA GTAGCTGAAT AAGATAGAGA ATTTTCTCT CAAAGAAAGT CTAAGTAGGC 780
	AGCTCAGAAG TAGTATGGCT GGAAGCAACC TGATGATATT GGGACCCCCA ACCTTCTTC 840
	GTCTTGTACC CATCATCCCC TAGTTGTTGA TCTCACTCAC ATAGTTGAAA ATCATCATAC 900
	TCCTGGGTT CATATCCAG TTATCAAGAA AGGGTCAAGA GAAGTCAGGC TCATTCCCTT 960
10	CAAAGACTCT AATTGGAAGT TAAACACATC AATCCCCCTC ATATTCCATT GACTAGAATT 1020
	TAATCACATG GCCACACCAA GTGCAAGGAA ATCTGGAAAA TATAATCTT ATTCCAGGTA 1080
	GCCATATGAC TCTTTAAAAT TCAGAAATAA TATATTTTA AAATATCATT CTGGCTTTGG 1140
	TATAAAGAAAT TGATGGTGTG GGGTGAGGAG GCCAAAATTA AGGTTGAGA GCCTATTATT 1200
	TTAGTTTAACTT CAAGAAATGA TGGTGTATG AATTAAAGGT A GACATAGGGG AGTGTGATG 1260
	AGGAGCTGTG AATGGATTAAAG AGAAACACTT GAGAGAATCA ATAGGACATG ATTTAGGGTT 1320
	GGATTGGAA AGGAGAAGAA AGTAGAAAGAG ATGATGCCTA CATTTCAC TTAGGCAATT 1380
15	TGTACCATTC AGTGAATAG GGAACACAGG AGGAAGAGCA GTGTTGGTG TATACAAAGA 1440
	GGAGGATGGA TGACGCATT CGTTTGGAT CTGAGATGTC TGTTGAACGT CCTAGTGGAG 1500
	ATGTCCACAA ACTCTTCTAC ATGTGGTTCT GAGTCAGGA CACAGATTTG GGCTGGAGAT 1560
	AGAGATATTG TAGGCTTATA CATAGAAATG GCATTTGAAT CTATAGAGAT AAAAAGACAC 1620
20	ATCAGAGGAA ATGTGTAAAG TGAGAGAGGA AAAGCCAAGT ACTGTGCTGG GGGGAATACC 1680
	TACATTAAA GGATGCAGTA GAAAGAAGCT AATAAACAAAC AGAGAGCAGA CTAACCAAAA 1740
	GGGGAGAAGA AAAACCAAGA GAATTCCACC GACTCCCAGG AGAGCATTTC AAGATTGAGG 1800
	GGATAGGTGT TGTGTTGAAT TTTGCAGCCT TGAGAATCAA GGGCCAGAAC ACAGCTTTA 1860
	GATTTAGCAA CAAGGAGTTT GGTGATCTCA GTGAAAGCAG CTTGATGGTG AAATGGAGGC 1920
	AGAGGAGAT TGCAATGAGT GAAACAGTGA ATGGGAAGTG AAGAAATGAT ACAGATAATT 1980
25	CTTGCTAAAA GCTTGGCTGT TAAAAGGAGG AGAGAAACAA GACTAGCTGC AAAGTGAGAT 2040
	TGGGTGTATG GAGCAGTTT AAATCTCAA AATAAGAGCT TTGTCCTTT TTGATTATGA 2100
	AAATAATGTG TTAATTGTA CTAATTGAGG CAATGAAAAA AGATAATAAT ATGAAAGATA 2160
	AAAATATAAA AACCAACCCAG AAATAATGAT AGCTACCATTT TTGATACAAT ATTTCTACAC 2220
	TCCTTCTAT GTATATATAC AGACACAGAA ATGCTTATAT TTTTATTAAA AGGGATTGTA 2280
	CTATACCTAA GCTGCTTTT CTAGTTAGTG ATATATATGG ACATCTCTCC ATGGCAACGA 2340
30	GTAATTGCAG TTATATTAAG TTCATGATAT TTCAACAATAA GGGCATATCT TTGCCCTTT 2400
	TATTTAATCA ATTCTTAATT GGTGAATGTT TGTTCCAGT TTGTTGTTGT TATTAACAAT 2460
	GTTCCCATAA GCATTCCTGT ACACCAATGT TCACACATT GTCTGATTTT TTCTTCAGGA 2520
	TAAAACCCAG GAGGTAGAAT TGCTGGTTG ATAGAAAGAGA AAGGATGATT GCCAAATTAA 2580
	AGCTTCAGTA GAGGGTACAT GCCGAGCACA AATGGGATCA GCCCTAGATA CCAGAAATGG 2640
	CACTTCTCA TTTCCCTGT GGACAAAAGG GAGAGAGGCA ATAACTGTGC TGCCAGAGTT 2700
35	AAATTGTCAGTA GTGGAGTAGC AGGAAATCAT TTGCTGAAAA TGAAAACAGA GATGATGTTG 2760
	TAGAGGTCT GAAGAGAGCA AAGAAAATTT GAAATTGCGG CTATCAGCTA TGGAAGAGAG 2820
	TGCTGAACTG GAAAACAAA GAAGTATTGA CAATTGGTAT GCTTGTAAATG GCACCGATT 2880
	GAACGCTGT GCCATTGTC ACCAGCAGCA CTCAGCAGCC AAGTTGGAG TTTTGTAGCA 2940
	GAAAGACAAA TAAGTTAGGG ATTTAATATC CTGGCCAAT GGTAGACAAA ATGAACTCTG 3000
40	AGATCCAGCT GCACAGGGAA GGAAGGGAG ACGGGAAGAG GTTAGATAGG AAATACAAGA 3060
	GTCAGGAGAC TGGAAAGATGT TGTGATATT AAGAACACAT AGAGTTGGAG TAAAAGTGT 3120
	AGAAAACATAG AAGGGTAAGA GACGGTCAG AAAGTAGGCT ATTTGAAGTT AACACTTCAG 3180
	AGGCAGAGTA GTTCTGAATG GTAACAAGAA ATTGAGTGTG CCTTGAGAG TAGGTTAAA 3240
	AACAATAGGC AACTTTATTG TAGCTACTTC TGGAACAGAA GATTGTCAATT AATAGTTTA 3300
	GAAAACAAA ATATATAGCA TACTTATTG TCAATTAAACA AAGAAAACATAT GTATTTTAA 3360
45	ATGAGATTAA ATGTTTATTG TAG 3383

(14) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11464 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human

(F) TISSUE TYPE:placenta

(iX) FEATURE:

5 (A) NAME/KEY: 5' UTR
 (B) LOCATION: 1..3
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 4..82
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 83..1453
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 1454..1465
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 1466..4848
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 4849..4865
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 4866..4983
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 4984..6317
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 6318..6451
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 6452..11224
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 11225..11443
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: 3' UTR
 (B) LOCATION: 11444..11464
 (C) IDENTIFICATION METHODS: E

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA
 48 Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala
 -35 -30 -25
 40 ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGG CTAATGCCAT
 98 Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
 -20 -15 -10
 45 AGAACAAATA CCAGGTTCAAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAAC
 158 158
 ATTAAGTGAC TCTTTGTGTC ACCAAATTTC ACTGTAATAT TAATGGCTCT TAAAAAAATA
 218
 GTGGACCTCT AGAAATTAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCACGT
 278
 GTCCTGGCAC TTTAATCAGC AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA
 338
 50 AAATCCCAAGT TTTCATGGGA AAATCCCAAGT TTTCATTGGA TTTCATGGG AAAATCCCA
 398
 GTACAAAATCTC GGGTGCATTC AGGAAATACA ATTTCCCAA GCAAATTGGC AAATTATGTA
 458
 AGAGATTCTC TAAATTTAGA GTTCCGTGAA TTACACCATT TTATGTAAAT ATGTTTGACA

518
 AGTAAAAAATT GATTCTTTT TTTTTTTCT GTGCCAGG CTGGAGTGCA GTGGCACAAAT
 578
 5 CTCTGCTCAC TGCAACCTCC ACCTCCTGGG TTCAAGCAAT TCTCCTGCCT CAGCCTCTG
 638
 AGTAGCTGGG ACTACAGGTG CATCCCGCCA TGCCTGGCTA ATTTTGGGT ATTTTTACTA
 698
 GAGACAGGGT TTTGGCATGT TGTCCAGGCT GGTCTTGGAC TCCTGATCTC AGATGATCCT
 758
 10 CCTGGCTCGG GCTCCAAAG TGCTGGGATT ACAGGCATGA ACCACACAC ATGGCCTAAA
 818
 AATTGATTCT TATGATTAAT CTCCTGTGAA CAATTGGCT TCATTTGAAA GTTGCCTTC
 878
 15 ATTTGAAACC TTCATTTAAA AGCCTGAGCA ACAAAAGTGAG ACCCCATCTC TACAAAAAAC
 938
 TGCAAAATAT CCTGTGGACA CCTCCTACCT TCTGTGGAGG CTGAAGCAGG AGGATCACTT
 998
 GAGCCTAGGA ATTTGAGCCT GCAGTGAGCT ATGATCCCAC CCCTACACTC CAGCCTGCAT
 1058
 GACAGTAGAC CCTGACACAC ACACACAAAA AAAAACCTTC ATAAAAAATT ATTAGTTGAC
 1118
 20 TTTTCTTAGG TGACTTTCCG TTTAAGCAAT AAATTAAAAA GTAAAATCTC TAATTTAGA
 1178
 AAATTATTAA TTAGTTACAT ATTGAAATT TTAAACCTTA GGTTAAGTT TTATGTCTAA
 1238
 ATTACCTGAG AACACACTAA GTCTGATAAG CTTCATTAA TGGGCCTTT GGATGATTAT
 1298
 25 ATAATATTCT GATGAAAGCC AAGACAGACC CTTAAACCAT AAAAATAGGA GTTCGAGAAA
 1358
 GAGGAGTAGC AAAAGTAAAAA GCTAGAATGA GATTGAATTC TGAGTCGAAA TACAAAATT
 1418
 TACATATTCT GTTTCTCTCT TTTTCCCCCT CTTAG CT GAA GAT GAT G GTAAA
 30 1470
 Ala Glu Asp Asp Glu
 -10
 GTAGAAATGA ATTTATTTT CTTTGCAAAC TAAGTATCTG CTTGAGACAC ATCTATCTCA
 1530
 35 CCATTGTCAG CTGAGGAAAA AAAAAATGG TTCTCATGCT ACCAATCTGC CTTCAAAGAA
 1590
 ATGTGGACTC AGTAGCACAG CTTGGAATG AAGATGATCA TAAGAGATAC AAAGAAGAAC
 1650
 CTCTAGCAAA AGATGCTTCT CTATGCCTTA AAAAATTCTC CAGCTCTTAG AATCTACAAA
 1710
 40 ATAGACTTTG CCTGTTTCAT TGGTCCTAAG ATTAGCATGA AGCCATGGAT TCTGTTGTAG
 1770
 GGGGAGCGTT GCATAGGAAA AAGGGATTGA AGCATTAGAA TTGTCCAAAA TCAGTAACAC
 1830
 CTCCTCTCAG AAATGCTTGT GGAAGAAGCC TGGAAAGGTTG CGGGTTGGTG GTGGGGTGGG
 1890
 45 GCAGAAAATT CTGGAAGTAG AGGAGATAGG AATGGGTGGG GCAAGAAGAC CACATTCA
 1950
 GCCAAAAGC TGAAAGAAC CATGGCATTG ATGATGAATT CAGGGTAATT CAGAATGGAA
 2010
 GTAGAGTAGG AGTAGGAGAC TGGTGAGAGG AGCTAGAGTG ATAAACAGGG TGTAGAGCAA
 2070
 50 GACGTTCTCT CACCCCAAGA TGTGAAATT GGACTTTATC TTGGAGATAA TAGGGTTAAT
 2130
 TAAGCACAAT ATGTATTAGC TAGGGTAAAG ATTAGTTGT TGTAAACAAAG ACATCCAAAG
 2190
 ATACAGTAGC TGAATAAGAT AGAGAATTAA TCTCTCAAAG AAAGTCTAAG TAGGCAGCTC
 2250

1 AGAAGTAGTA TGGCTGGAAG CAACCTGATG ATATTGGAC CCCCCAACCTT CTTCAGTCCT
 2310
 2 GTACCCATCA TCCCCTAGTT GTTGATCTCA CTCACATAGT TGAAAATCAT CATACTTCCT
 2370
 3 GGGTTCATAT CCCAGTTATC AAGAAAGGGT CAAGAGAAGT CAGGCTCATT CCTTTCAAAG
 2430
 4 ACTCTAATTG GAAGTTAAC ACATCAATCC CCCTCATATT CCATTGACTA GAATTAAATC
 2490
 5 ACATGGCCAC ACCAAGTGCA AGGAAATCTG GAAAATATAA TCTTATTCC AGGTAGCCAT
 2550
 6 ATGACTCTT AAAATTAGA AATAATATAT TTTTAAAATA TCATTCTGGC TTTGGTATAA
 2610
 7 AGAATTGATG GTGTGGGTG AGGAGGCCAA AATTAAGGGT TGAGAGCCTA TTATTTAGT
 2670
 8 TATTACAAGA AATGATGGTG TCATGAATTA AGGTAGACAT AGGGAGTGC TGATGAGGAG
 2730
 9 CTGTGAATGG ATTTTAGAAA CACTTGAGAG AATCAATAGG ACATGATTAA GGGTTGGATT
 2790
 10 TGGAAAGGAG AAGAAAGTAG AAAAGATGAT GCCTACATTTC TTCACCTAGG CAATTGTAC
 2850
 11 CATTCACTGA AATAGGGAAC ACAGGAGGAA GAGCAGGTTT TGGTGTATAC AAAGAGGAGG
 2910
 12 ATGGATGACG CATTTCGTTT TGGATCTGAG ATGTCTGTGG AACGTCTAG TGGAGATGTC
 2970
 13 CACAAACTCT TCTACATGTG GTTCTGAGTT CAGGACACAG ATTTGGCCTG GAGATAGAGA
 3030
 14 TATTGTAGGC TTATACATAG AAATGGCATT TGAATCTATA GAGATAAAA GACACATCAG
 3090
 15 AGGAAATGTG TAAAGTGAGA GAGGAAAAGC CAAGTACTGT GCTGGGGGAA ATACCTACAT
 3150
 16 TTAAAGGATG CAGTAGAAAG AAGCTAATAA ACAACAGAGA GCAGACTAAC CAAAAGGGGA
 3210
 17 30 GAAGAAAAAC CAAGAGAATT CCACCGACTC CCAGGAGAGC ATTTCAAGAT TGAGGGATA
 3270
 18 GGTGTTGTGT TGAATTTGAGA AGCCTTGAGA ATCAAGGGCC AGAACACAGC TTTTAGATT
 3330
 19 AGCAACAAAG AGTTGGTGA TCTCAGTGAAGC AGCAGCTTGA TGGTGAATG GAGGCAGAGG
 3390
 20 CAGATTGCAA TGAGTGAAC AGTGAATGGG AAGTGAAGAA ATGATACAGA TAATTCTTGC
 3450
 21 TAAAAGCTTG GCTGTTAAAA GGAGGAGAGA AACAAAGACTA GCTGCAAAGT GAGATTGGGT
 3510
 22 TGATGGAGCA GTTTAAATC TCAAAATAAA GAGCTTGTG CTTTTTGAT TATGAAAATA
 3570
 23 40 ATGTGTTAAT TGTAACATAAT TGAGGCAATG AAAAAAGATA ATAATATGAA AGATAAAAAT
 3630
 24 ATAAAAACCA CCCAGAAATA ATGATAGCTA CCATTTGAT ACAATATTC TACACTCCCT
 3690
 25 TCTATGTATA TATACAGACA CAGAAATGCT TATATTTTA TAAAGGGAA TTGTACTATA
 3750
 26 CCTAAGCTGC TTTTCTAGT TAGTGTATA TATGGACATC TCTCCATGGC AACGAGTAAT
 3810
 27 TGCAAGTATA TTAAGTCAT GATATTCAAC AATAAGGGCA TATCTTGCC CTTTTTATT
 3870
 28 AATCAATTCT TAATTGGTGA ATGTTGTTT CCAGTTGTT GTTGTATTAA ACAATGTTCC
 3930
 29 CATAAGCATT CCTGTACACC AATGTTACA CATTGTCTG ATTTTTCTT CAGGATAAAA
 3990
 30 CCCAGGAGGT AGAATTGCTG GGTTGATAGA AGAGAAAGGA TGATTGCCAA ATTAAAGCTT
 4050
 31 CAGTAGAGGG TACATGCCGA GCACAAATGG GATCAGCCCT AGATACCAGA AATGGCACTT

4110 TCTCATTCC CCTTGGGACA AAAGGGAGAG AGGCAATAAC TGTGCTGCCA GAGTTAAATT
 4170 5 TGTACGTGGA GTAGCAGGAA ATCATTGCT GAAAATGAAA ACAGAGATGA TGTTGTAGAG
 4230 GTCCTGAAGA GAGCAAAGAA AATTGAAAT TGCGGCTATC AGCTATGGAA GAGAGTGCTG
 4290 10 AACTGGAAAA CAAAAGAAGT ATTGACAATT GGTATGCTTG TAATGGCACC GATTTGAACG
 4350 CTTGTGCCAT TGTTCACCAAG CAGCACTCAG CAGCCAAGTT TGGAGTTTG TAGCAGAAAG
 4410 15 ACAAAATAAGT TAGGGATTAA ATATCCTGGC CAAATGGTAG ACAAAATGAA CTCTGAGATC
 4470 CAGCTGCACA GGGAGGAAG GGAAGACGGG AAGAGGTTAG ATAGGAAATA CAAGAGTCAG
 4530 20 GAGACTGGAA GATGTTGTGA TATTAAAGAA CACATAGAGT TGGAGTAAA GTGTAAGAAA
 4590 ACTAGAAGGG TAAGAGACCG GTCAGAAAGT AGGCTATTG AAGTTAACAC TTCAGAGGCA
 4650 25 GAGTAGTTCT GAATGGTAAC AAGAAATTGA GTGTGCCTT GAGAGTAGGT TAAAAAAACAA
 4710 TAGGCAACTT TATTGTAGCT ACTTCTGGAA CAGAAGATTG TCATTAATAG TTTAGAAAA
 4770 30 CTAAAATATA TAGCATACTT ATTTGTCAAT TAACAAAGAA ACTATGTATT TTTAAATGAG
 4830 ATTTAATGTT TATTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT
 4880 35 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu
 -5 1 5
 GAA TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC
 4928 40 Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe
 10 15 20
 ATT GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC
 4976 45 Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp
 25 30 35
 TGT AGA G GTATTTTT TTAATTCGCA AACATAGAAA TGACTAGCTA CTTCTTCCCA
 5032 50 Cys Arg Asp
 40 TTCTGTTTA CTGCTTACAT TGTTCCGTGC TAGTCCCAAT CCTCAGATGA AAAGTCACAG
 5092 55 GAGTGACAAT AATTCACTT ACAGGAAACT TTATAAGGCA TCCACGTTT TTAGTTGGGG
 5152 60 TAAAAAAATTG GATACAATAA GACATTGCTA GGGGTCAATGC CTCTCTGAGC CTGCCTTGA
 5212 65 ATCACCAATC CCTTTATTGT GATTGCATTA ACTGTTAAA ACCTCTATAG TTGGATGCTT
 5272 70 AATCCCTGCT TGTTACAGCT GAAAATGCTG ATAGTTACC AGGTGTGGTG GCATCTATCT
 5332 75 GTAATCCTAG CTACTTGGGA GGCTCAAGCA GGAGGATTGC TTGAGGCCAG GACTTTGAGG
 5392 80 CTGTAGTACA CTGTGATCGT ACCTGTGAAT AGCCACTGCA CTCCAGCCTG GGTGATATAC
 5452 85 AGACCTTGTC TCTAAAATTA AAAAAAAA AAAAAAAAC CTTAGGAAAG GAAATTGATC
 5512 90 AAGTCTACTG TGCCCTTCAA AACATGAATT CCAAATATCA AAGTTAGGCT GAGTTGAAGC
 5572 95 AGTGAATGTG CATTCTTTAA AAATACTGAA TACTTACCTT AACATATATT TTAAATATTT
 5632

TATTTAGCAT TTAAAAGTTA AAAACAATCT TTTAGAATTG ATATCTTAA AATACTCAA
 5692
 AAAGTTGCAG CGTGTGTGTT GTAATACACA TTAAACTGTG GGGTTGTTG TTTGTTGAG
 5752
 5 ATGCAGTTTC ACTCTGTCAC CCAGGCTGAA GTGCAGTGCA GTGCAGTGGT GTGATCTCGG
 5812
 CTCACTACAA CCTCCACCTC CCACGTTCAA GCGATTCTCA TGCCCTAGTC TCCCGAGTAG
 5872
 GTGGGATTAC AGGCATGCAC CACTTACACC CGGCTAATT TTGTATTTT AGTAGAGCTG
 5932
 GGGTTTCAAC ATGTTGGCCA GGCTGGTCTC AAACCCCTAA CCTCAAGTGA TCTGCCTGCC
 5992
 TCAGCCTCCC AAACAAACAA ACAACCCAC AGTTAATAT GTGTTACAAC ACACATGCTG
 6052
 15 CAACTTTAT GAGTATTTA ATGATATAGA TTATAAAAGG TTGTTTTAA CTTTTAAATG
 6112
 CTGGGATTAC AGGCATGAGC CACTGTGCCA GCCCTGAACG GTGTTTTAA AAATGTCTGA
 6172
 CCAGCTGTAC ATAGTCTCCT GCAGACTGGC CAAGTCTCAA AGTGGGAACA GGTGTATTAA
 6232
 20 GGACTATCCT TTGGTTAAAT TTCCGCAAAT GTTCCGTGCA AAGAATTCTT CTAACCTAGAG
 6292
 TTCTCATTTA TTATATTTAT TTCAG AT AAT GCA CCC CGG ACC ATA TTT ATT
 6343

Asp Asn Ala Pro Arg Thr Ile Phe Ile
 40 45

25 ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC
 6391
 Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
 50 55 60
 TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT
 6439
 30 Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile
 65 70 75 80
 ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTTT CAATCATGTT AATATAATCA
 6496
 Ile Ser Phe Lys
 ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA AGTAATGTAA TTAGAAAATC
 6556
 35 CAAATATCCT CAGACCAACC TTTTGTCTAG AACAGAAATA ACAAGAAGCA GAGAACATT
 6616
 AAAGTGAATA CTTACTAAAA ATTATCAAAC TCTTTACCTA TTGTGATAAT GATGGTTTT
 6676
 40 CTGAGCCTGT CACAGGGAA GAGGAGATAC AACACTGTT TTATGACCTG CATCTCCTGA
 6736
 ACAATCAGTC TTTATACAAA TAATAATGTA GAATACATAT GTGAGTTATA CATTAAAGAA
 6796
 TAACATGTGA CTTTCCAGAA TGAGTTCTGC TATGAAGAAT GAAGCTAATT ATCCTTCTAT
 6856
 45 ATTTCTACAC CTTTGTAAAT TATGATAATA TTTTAATCCC TAGTTGTTT GTTGCTGATC
 6916
 CTTAGCCTAA GTCTTAGACA CAAGCTTCAG CTTCCAGTTG ATGTATGTTA TTTTTAATGT
 6976
 TAATCTAATT GAATAAAAGT TATGAGATCA GCTGAAAAG TAATGCTATA ATTATCTCA
 7036
 50 AGCCAGGTAT AAAGTATTTC TGGCCTCTAC TTTTCTCTA TTATTCTCCA TTATTATTCT
 7096
 CTATTATTTC TCTCTATTTC CTCCATTATT GTTAGATAAA CCACAATTAA CTATAGCTAC
 7156
 AGACTGAGCC AGTAAGAGTA GCCAGGGATG CTTACAAATT GGCAATGCTT CAGAGGAGAA
 7216

55

TTCCATGTCA TGAAGACTCT TTTTGAGTGG AGATTTGCCA ATAAATATCC GCTTTCATGC
 7276
 CCACCCAGTC CCCACTGAAA GACAGTTAGG ATATGACCTT AGTGAAGGTA CCAAGGGGCA
 7336
 5 ACTTGGTAGG GAGAAAAAAAG CCACTCTAAA ATATAATCCA AGTAAGAACCA GTGCATATGC
 7396
 AACAGATACA GCCCCCAGAC AAATCCCTCA GCTATCTCCC TCCAACCAGA GTGCCACCCC
 7456
 10 TTCAGGTGAC AATTGGAGT CCCCATTCTA GACCTGACAG GCAGCTTAGT TATCAAAATA
 7516
 GCATAAGAGG CCTGGGATGG AAGGGTAGGG TGAAAGGGT TAAGCATGCT GTTACTGAAC
 7576
 AACATAATTA GAAGGGAAAGG AGATGGCAA GCTCAAGCTA TGTGGATAG AGGAAAACTC
 7636
 15 AGCTGCAGAG GCAGATTCAAG AAACTGGAT AAGTCCGAAC CTACAGGTGG ATTCTTGTG
 7696
 AGGGAGACTG GTGAAAATGT TAAGAAGATG GAAATAATGC TTGGCACTTA GTAGGAAC
 7756
 GGCAAATCCA TATTGGGGGG AGCCTGAAGT TTATTCAATT TTGATGGCCC TTTTAAATAA
 7816
 20 AAAGAATGTG GCTGGCGTG GTGGCTACA CCTGTAATCC CAGCACTTG GGAGGCCGAG
 7876
 GGGGCGGAT CACCTGAAGT CAGGAGTTCA AGACCAGCCT GACCAACATG GAGAAACCCC
 7936
 ATCTCTACTA AAAATACAAA ATTAGCTGGG CGTGGTGGCA TATGCCGTGTA ATCCCAGCTA
 7996
 25 CTCGGGAGGC TGAGGCAGGA GAATCTTTG AACCCGGGAG GCAGAGGTG CGATGAGCCT
 8056
 AGATCGTGCC ATTGCACTCC AGCCTGGCA ACAAGAGCAA AACTCGGTCT CAAAAAAA
 8116
 AAAAAAAAAG TGAAATTAAC CAAAGGCATT AGCTTAATAA TTTAATACTG TTTTAAGTA
 8176
 30 GGGCGGGGGG TGGCTGGAAG AGATCTGTGT AAATGAGGGA ATCTGACATT TAAGCTTCAT
 8236
 CAGCATCATA GCAAATCTGC TTCTGGAAGG AACTCAATAA ATATTAGTTG GAGGGGGGA
 8296
 GAGAGTGAGG GGTGGACTAG GACCAGTTT AGCCCTTGTC TTTAATCCCT TTTCTGCCA
 8356
 35 CTAATAAGGA TCTTAGCAGT GGTTATAAAA GTGGCCTAGG TTCTAGATAA TAAGATACAA
 8416
 CAGGCCAGGC ACAGTGGCTC ATGCCCTATAA TCCCAGCACT TTGGGAGGGC AAGGCGAGTG
 8476
 TCTCACTTGA GATCAGGAGT TCAAGACCAG CCTGGCCAGC ATGGCGATAC TCTGTCTCTA
 8536
 40 CTAAAAAAA TACAAAATT AGCCAGGCAT GGTGGCATGC ACCTGTAATC CCAGCTACTC
 8596
 GTGAGCCTGA GGCAGAAGAA TCGCTTGAAA CCAGGAGGTG TAGGCTGCAG TGAGCTGAGA
 8656
 TCGCACCACT GCACTCCAGC CTGGGGACAA GAATGAGACT TTGTCTAAA AAAAGAAAAA
 8716
 45 GATACAACAG GCTACCCCTTA TGTGCTCACC TTTCACTGTT GATTACTAGC TATAAAGTCC
 8776
 TATAAAAGTTC TTTGGTCAAG AACCTTGACA ACACAAAGAG GGATTTGCTT TGAGAGGTTA
 8836
 50 CTGTCAGAGT CTGTTTCATA TATATACATA TACATGTATA TATGTATCTA TATCCAGGCT
 8896
 TGGCCAGGGT TCCCTCAGAC TTTCCAGTGC ACTTGGGAGA TGTTAGGTCA ATATCAACTT
 8956
 TCCCTGGATT CAGATTCAAC CCCTTCTGAT GTAAAAAAA AAAAAAAA GAAAGAAATC
 9016
 CCTTCCCT TGGAGCACTC AAGTTTCACC AGGTGGGCT TTCCAAGTTG GGGTTCTCC

9076 AAGGTCAATTG GGATTGCTT CACATCCATT TGCTATGTAC CTTCCCTATG ATGGCTGGGA
 9136
 5 GTGGTCAACA TCAAAACTAG GAAAGCTACT GCCCAAGGAT GTCCTTACCT CTATTCTGAA
 9196
 ATGTGCAATA AGTGTGATTA AAGAGATTGC CTGTTCTACC TATCCACACT CTCGCTTCA
 9256
 ACTGTAACTT TCTTTTTTC TTTTTCTT TTTTTCTTT TTTTGAAAC GGAGTCTCGC
 9316
 10 TCTGTCGCC AGGCTAGAGT GCAGTGGCAC GATCTCAGCT CACTGCAAGC TCTGCCTCCC
 9376
 GGGTTCACGC CATTCTCCTG CCTCACCCCTC CCAAGCAGCT GGGACTACAG GCGCCTGCCA
 9436
 CCATGCCAG CTAATTTTT GTATTTTAG TAGAGACGGG GTTCACCGT GTTAGCCAGG
 9496
 15 ATGGTCTCGA TCTCCTGAAC TTGTGATCCG CCCGCCTCAG CCTCCCAAAG TGCTGGGATT
 9556
 ACAGGCGTGA GCCATCGCAC CCGGCTCAAC TGTAACCTTC TATACTGGTT CATCTTCCCC
 9616
 TGTAATGTTA CTAGAGCTTT TGAAGTTTG GCTATGGATT ATTTCTCATT TATACATTAG
 9676
 20 ATTTCAGATT AGTTCCAAAT TGATGCCAC AGCTTAGGGT CTCTCCTAA ATTGTATATT
 9736
 GTAGACAGCT GCAGAAGTGG GTGCCAATAG GGGAACTAGT TTATACTTTC ATCAACTTAG
 9796
 GACCCACACT TGTTGATAAA GAACAAAGGT CAAGAGTTAT GACTACTGAT TCCACAACTG
 9856
 25 ATTGAGAAGT TGGAGATAAC CCCGTGACCT CTGCCATCCA GAGTCTTCA GGCATTTG
 9916
 AAGGATGAAG AAATGCTATT TTAATTTGG AGGTTTCTCT ATCAGTGCTT AGGATCATGG
 9976
 GAATCTGTGC TGCCATGAGG CCAAAATTAA GTCCAAAACA TCTACTGGTT CCAGGATTA
 10036
 30 CATGGAAGAA CCTTAGGTGG TGCCCACATG TTCTGATCCA TCCTGAAAA TAGACATGCT
 10096
 GCACTAACAG GAAAAGTGCA GGCAGCACTA CCAGTTGGAT AACCTGCAAG ATTATAGTT
 10156
 CAAGTAATCT AACCAATTCT CACAAGGCC TATTCTGTGA CTGAAACATA CAAGAATCTG
 10216
 35 CATTGGCCT TCTAAGGCAG GGCCCAGCCA AGGAGACCAT ATTCAGGACA GAAATTCAAG
 10276
 ACTACTATGG AACTGGAGTG CTTGGCAGGG AAGACAGAGT CAAGGACTGC CAACTGAGCC
 10336
 AATACAGCAG GCTTACACAG GAACCCAGGG CCTAGCCCTA CAACAATTAT TGGGTCTATT
 10396
 40 CACTGTAAGT TTTAATTCA GGCTCCACTG AAAGAGTAAG CTAAGATTCC TGGCACTTTC
 10456
 TGTCTCTCTC ACAGTTGGCT CAGAAATGAG AACTGGTCAG GCCAGGCATG GTGGCTTACA
 10516
 45 CCTGGAATCC CAGCACTTG GGAGGCCGAA GTGGGAGGGT CACTTGAGGC CAGGAGTTCA
 10576
 GGACCACTT AGGCAACAAA GTGAGATACC CCCTGACCCC TTCTCTACAA AAATAAATT
 10636
 TAAAAAATTAG CCAAATGTGG TGGTGTATAC TTACAGTCCC AGCTACTCAG GAGGCTGAGG
 10696
 50 CAGGGGGATT GCTTGAGCCC AGGAATTCAA GGCTGCAGTG AGCTATGATT TCACCACTGC
 10756
 ACTTCTGGCT GGGCAACAGA GCGAGACCT GTCTCAAAGC AAAAAGAAAA AGAAACTAGA
 10816
 ACTAGCCTAA GTTTGTGGGA GGAGGTCATC ATCGTCTTTA GCCGTGAATG GTTATTATAG
 10876

AGGACAGAAA TTGACATTAG CCCAAAAAGC TTGTGGTCTT TGCTGGAAC TCACTTAATC
 10936
 TTGAGCAAAT GTGGACACCA CTCAATGGGA GAGGAGAGAA GTAAGCTGTT TGATGTATAG
 10996
 5 GGGAAAACTA GAGGCCTGGA ACTGAATATG CATCCCATGA CAGGGAGAAT AGGAGATTG
 11056
 GAGTTAAGAA GGAGAGGAGG TCAGTACTGC TGTCAGAGA TTTTTTTAT GTAACCTTG
 11116
 10 AGAAGCAAAA CTACTTTGT TCTGTTGGT AATATACTTC AAAACAAACT TCATATATTC
 11176
 AAATTGTTCA TGTCCCTGAAA TAATTAGGTA ATGTTTTTT CTCTATAG GAA ATG AAT
 11233

Glu Met Asn
 85

15 CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG
 11281
 Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Glu
 90 95 100
 AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA
 11329

20 Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser
 105 110 115
 TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA
 11377
 Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys
 120 125 130 135

25 CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC
 11425
 Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe
 140 145 150

30 ACT GTT CAA AAC GAA GAC TAGCTATTAA AATTTCATGC C
 11464
 Thr Val Gln Asn Glu Asp
 155

(15) INFORMATION FOR SEQ ID NO:14:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28994 base pairs
 (B) TYPE:nucleic acid
 (C) STRANDEDNESS:double
 (D) TOPOLOGY:linear

40 (ii) MOLECULE TYPE:Genomic DNA
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM:human
 (F) TISSUE TYPE:placenta

45 (ix) FEATURE:
 (A) NAME/KEY: 5' UTR
 (B) LOCATION: 1..15606
 (C) IDENTIFICATION METHODS:E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 15607..15685
 (C) IDENTIFICATION METHODS:S
 (A) NAME/KEY: intron
 (B) LOCATION: 15686..17056
 (C) IDENTIFICATION METHODS:E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 17057..17068
 (C) IDENTIFICATION METHODS:S
 (A) NAME/KEY: intron
 (B) LOCATION: 17069..20451

5 (C) IDENTIFICATION METHODS:E
 (A) NAME/KEY:leader peptide
 (B) LOCATION:20452..20468
 (C) IDENTIFICATION METHODS:S
 (A) NAME/KEY:mat peptide
 (B) LOCATION:20469..20586
 (C) IDENTIFICATION METHODS:S
 (A) NAME/KEY:intron
 (B) LOCATION:20587..21920
 (C) IDENTIFICATION METHODS:E
 10 (A) NAME/KEY:mat peptide
 (B) LOCATION:21921..22054
 (C) IDENTIFICATION METHODS:S
 (A) NAME/KEY:intron
 (B) LOCATION:22055..26827
 (C) IDENTIFICATION METHODS:E
 (A) NAME/KEY:mat peptide
 (B) LOCATION:26828..27046
 (C) IDENTIFICATION METHODS:S
 15 (A) NAME/KEY:3'UTR
 (B) LOCATION:27047..28994
 (C) IDENTIFICATION METHODS:E
 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

25 ACTTGCCTTA AAAGCTTGAC ATAGGTAGAC AACATTAGAT TAATTCCTT GCTCACATCT
 60
 25 GTTCAAGAAA AATCATTAA GTTATAAAAT ATAACAAACC TTCTGCATTA TAAGACTGAT
 120
 25 GTTTAGAAAT ATAAACATTT TATACATCAC CATTAAATC TTTCTCCAAG GCTTCATCTT
 180
 25 TATAAAATAG TCCGGAAATT TCAGAGAAAG ATGAATCTGA TTTTCCAAGA GAGGACAGCT
 240
 30 GTGGACTATC TGGCACTGGA GACTAAATAA AGAAAGCAGG TACAGTCAT AAGATCTTCA
 300
 30 GGACATATAC ATTTTGTAA TTAAGAAAAA GCAAATAAAA CATTTCAG AAAAAGGCAA
 360
 35 ACATGCTAGA AAGCATATGA CTTAGTCATT TGAGTTTTA TTATTAAGGA AATTACAGG
 420
 35 CCCAAGAAC ACCTTGCTCA ATATATTAAA TTTTATTTG GTTTCAACT AGACTTGCT
 480
 40 TTTCATTGT TTGTTTTGT GACAAGTTCT CGCTCTGTCA CCTAGGCCAA AGTGTAGTGA
 540
 40 CACAATCTTA GCTCACTGTA GCCTCCTAGA TTCAAGTGAT CCTCCTGTCT CAGACTCCTG
 600
 45 AGTAGCTAGG ACTACAGGAA CATTCCACCA TGCCCAGCTA ATTTGTTTT GTTTGTTTT
 660
 45 GTTTCAAGAG ACAATGTATT GCAGCGTTGC CCAGGCTGAT CTGAAACTCT TAGCCTCAAA
 720
 45 CGATACTCCT GCCTCAGCCT CCCAAAGCAC TAGGATTACA GACATGAGCC AATGCGCCCA
 780
 50 GCCTTAAATT AGACTTTAAA TGTGGTTTA AACTCCTGTT GAAAAAGCGT CTGGTATCTT
 840
 50 GAACCACTAG ATGTTTCAT AGCAATGAAG CTAAACTGTA ATTTAGACAG TAGCCAAATG
 900
 55 CTTGTGAAAT TTTGCTAAAT AATATAATCT TCAAGGGAGC AAATCATGTC CCAAATGCAA
 960
 55 AAGATCAACT GGTGGGGGCA GTAGTAAAAG ACAGGATACT GTGCTCTTA AAAGGTCAGT
 1020
 55 AACTATAGTA CCTAGTTATC TTACTTATCA CAGCAAAATA ATTACATAAA ATCCTATGGA
 1080
 55 TCATAAAAGGC ACAGACTCAC TTCTGTCTCT AGATCTCAAG CTACCAAAAA GAAATCTCCC

1140
 AATAGTTCT TGGAGGCCTA TACTTAGTGA AAAAGCAGCT GGAATCAACA TAGTTCCTCC
 1200
 5 TATGTTGTAG GACAATCCTA GCTCTGGCA TACGAATACA TTAAATCCCA CTTATCTATA
 1260
 GAGCTTCCTT AAAGGGAAGA AATTTGAGTA GTATGTAAAA CAGAATAAAA GATTAAGGCT
 1320
 10 CCATAGGCAT ACAGCTTACC TCCAATTCTC TTGGCCTCTT GCAATTCTA TTATCAGGCT
 1380
 TTACAAGGTG ATTTGCCATC ATATTCCGAA GGCACCAGCT ACAAGCTTA GAACAATGCC
 1440
 15 AGATTTAGGT ACAAACTCCA TGCTACAAGC TCTCTGGAAT CCTTCCCTGT TTCCCACCTCC
 1500
 TACTGCTGAT GTTAATTAG ACTGTCATTA TCTGTCACTT TCCTAAACTC AATTTCTCCC
 1560
 20 TCCTCTAAAT CATTCTATCA ACTGCTATTG GGGTAATCTT TCAAAACTTT GATTACTGCA
 1620
 TTCCTTAAAC TCAAAAACCT TCATTGTTCC AGAATAAGTT GAAATTCCAT GATATGGCCT
 1680
 TCAAGGTCT GTATTATCTG GTGCAAGCCT ACTAGTCCCA TCATTTCAA CTACTCCTCT
 1740
 25 CTATGTACTT AGCCAAATGA GTCTCTCTGG CAATTCTGCC TTGTTTCAGG ACTGGCTCAG
 1800
 TTAAGATTCT TTTATCTTCG GCCGGGCGCG CTGGCTCACG GCTGTAATCC CAGCACTTG
 1860
 GGAAGCTGAG GCAGGAAGAT CACCTGAGGT CGGGAGTTCG AGACCAGCCT GGCCAGCATG
 1920
 30 GTGAAACCT GTGTCTACTA AAAATCCAAA CATTAGCCAG GCGTGGTGGC AGGCGCTGT
 1980
 AATCCCAGCT ACTTGGGAAG CTGAGGTGAG AGAATCGCTT GAACCCAGGA GAGGGAGGTT
 2040
 GCAGTGAGCC GAGATTGTGC CATTGCACTC CAGCCTGGC AACAGAGCGA GACTCCACCT
 2100
 35 CAAAAAAA AAGGATTCTT CTATCTTCAC AAAATCTAA TGTTAAACA GGTCTTACAG
 2160
 TTCATCTAAT TCAATCTCAT TTTTACAAG TGAGAAAACA GGGACAGTGA CGGTGGATCA
 2220
 AGTGACACCA GTAAGACTGA GCTAAATTAG AACCGAGATC TCACTCGAGT CTGAGGTTAT
 2280
 40 TCCCACGTGC CAACCTTAAT TTAAAGTAGC TTCAAATTCTT ACTTTACTT TTCCATAAAT
 2340
 TCGGAAGGGA TTTCCCTAG GAGTCCAAAT GTGAAACCT GGAAGGGTAT AGTCTCTGTG
 2400
 TCTTGAGAT GAGGGGAGCC CTGTCCATAT TCAAGTTATC AATTGACTTT GTTGTGTTTG
 2460
 45 AGAAACGATG CTGATTTGGG TAACTTTAAC ACATCTGTT GATTAGTCCT ATAAAATATG
 2520
 CATATATAGA AGACAGAAAG AGCAACAACA AATTGAAAG ATGCTTGTAA AGTAAATTCT
 2580
 GTATCGTACG TGTCCATTCC TGCCAGTACC TTTATAGTAT GTAGTTAC GTGCTGTAAT
 2640
 AGTATTAATA GTATCTAGAA AATACTACAC ATGCACAGCA GTGCTAACTT TGCCTTGGGA
 2700
 50 GTTGGAAAAT ACTTCAGAGA AGCCAACAGG CAGATTTTC TCTCTCCCT TCCCCTCTA
 2760
 ATTTCCCTT TCCCCTTCAC CCCCTTCTCT TCTCTCCCCA AGTAACACTG TGCACCTATG
 2820
 TCAAACGAAA ACTTATAATC AAGTAACTGT TTCTGAAAA ATAAGTTCGT TTTCTGTCA
 2880
 55 TGGCTCAAGG CCTCAGCAGA TCCAGGCCTG GTGGACGGGC TGGTCTTCGT CGTGTGCCAA
 2940

ACACTGACCA CTGCCCTGGC TCTGCCATCT TAGGCTTAGT GACCTGGCTG TTACTAAGCA
 3000
 CTGTCCCCCTC TGCCCCATGC AGCTGTCTCC TTCTAGTCTT CTCCCTCTTC TCAACGCGAT
 3060
 5 CCTAGCCCCCT CAGGCCATTT CACCTCCATT TTCCCTCACT TCCCGCCGCC CCTCCGCACT
 3120
 TCCTCCCTAC TGTTGTTCC GCCCCACTAG AGCCCCTCAG AGAAAGTTTC CATCCTCGCA
 3180
 10 CCCTTCCTTG TGTCACAGCC CGTCACATTC TCACAGGCGC CCATCCCTCC AGCCCCACCC
 3240
 CAAGGCCAAT GTACTTCGCG GTATGGGGAC CTTCCTCGTC AGCGAACGCG AGGGAGTGAA
 3300
 GACCCCTGGGC GCGGGGTGCT CGGACTTCGG GGGTGGAGGT GGGAAAGCGCG CCGCACTCCC
 3360
 15 AGCAGCCCCCT GCACGAGTCA CGTGACAGCT CTCCCACAC CACCCCCCCCC AACTCCCCA
 3420
 CCGTAGCCTC CCAGAGCCAG GCCCCACGGA AAGGCAGCTT TTTCCCGGTT TTCTCCCGCT
 3480
 CTTTCCCCCTC CACTTGGAAAT ACTCGTGAAA CAAAAATCTC TCCCTGCCAC CCTGTGTGTG
 3540
 20 TTTGAACCAAG GAAAAAAATCT GAAACTGGTC AAGAAAGAAC AAGGAAGACT TGCCAAAGCA
 3600
 AGGCCGGTGT GTGTCCCAGC AGCTTAGAAT CTCAGCAAAG GAACACAAAAA TAGCACATCC
 3660
 ACGGCCTCTT TTCGAGTAAA ATTACTTGG TTTGTTGCA GGAAGGGTTT AAAACTGCGT
 3720
 25 TTGCAGATGC TCTGTTGCA GGAAGGCTTT AATCACGTGT TCCCGTGGCC CACAAGCAAG
 3780
 GCTTTAGAT CCAGAGCCTC AGTTACTGCC CCCTCTTCCT CTTTGGTGCA ACCAAACGTT
 3840
 CAGAACACG CCTTCTTAGA AAATTCTTAC CCCGGGTGTG TCAATAAGTT AAGTCTAATT
 3900
 30 GGCAACAGCT ATCAAAAAGT GTTGCATAAC ACACATGGCT CACATAATTG TAGCTTGCC
 3960
 TCATCGGGTG TTTTAATGCG GAGGCTTTGA CCTGCAATT CAAAGATATA CATTCCAAGC
 4020
 TTACGCCAG TTAGTGGATG TGGAAAGAAAA AAAAAAGCAA ATTACCTCAT AACACAAAGG
 4080
 35 TCAATAACAC ACATCCATAA GCTCCAGGTA CAAAATCTTA CATCTTAGAG AACTATATT
 4140
 AACATTTACA TACATTACTA AGGTTTTTT TTTCTTTTG CTTGATTAAA TGTTAGTTAT
 4200
 40 CATTAAGTCT TGGAATTATT CTGTGTGTGT ATATTTATTT GCTGTTGTG AAGAAGCCGG
 4260
 TTGTTTTAAA TAAGTTCTA GAAAATAAGC GCTCAATGTG TTTAATCTGA GTTGCTAATA
 4320
 TTGTGAAATA TAGGCCACAT AATACTAGCC TAGATAACTA TGGCGAAGTA AGGAGTCTCA
 4380
 45 AACACTGTCC CAGAACAAATA GCAATCTGTG TTGAATTTTT ACCCTCTGTG GTAAAATGAA
 4440
 GGGAAAAGGA ATGAAGTTTT AGTTGCCTT AATTTTTATC TTTATTGTT CAGACTCTTC
 4500
 AGCAGTATAA AGTTTCATC AAGTCAAATA TATTCACTTT AAAGTGAATG TGCTTTATTC
 4560
 50 TGATACCATG TCCTTCCTAA TTTGGGGGGC CAGGTGAGAT AAGTTTTATG AAATAAAAAG
 4620
 ATTAAAAATT CTTACATTT TAGTGTCTT CCTTGGTAAA ATGTAGAGTT GTCCACTGTG
 4680
 TTTATCTCCT CCTCCTTATT ATCATGGTTG CTGTTATTAT TTTTAATGGT TCATTAAACC
 4740
 55 CAAGGGTCTG GGAAATACTC ATGGAATTCA TCTCACAGCC TTCACACTGT ATGATATTAA

4800
 AACAGGTGGT TGTCCATCTG ATTCTTAAAA TATTTCCAAG AAAAATGATT CCACCTAAT
 4860
 5 CATAAAATGCT TTCATCAGAT TAAGAGAACCA CCATGGACAT TTTATTTTAT TTTATTTT
 4920
 AAATATTAAC TTCCATTGCA TAAGCTAAAT GGGTAGGAAT AAGTGAGATG ATATTGTTA
 4980
 CTAGAGCTTT AAAATATTCA AAGGGCTGTC ATCATTATCT CATTAAATCT TTGAAAACA
 5040
 10 CTCTATGAAG TACAAAGGAC ACTGAGACAT TTGTTGCTCT ATATCAAAGA AAAAAGTGT
 5100
 TGTCCCAAAA CTTCAAAATG TGTAAATTAC ACATTCTGCA TCTTACAGC TGGAGAAA
 5160
 15 TCACTGGCAA TGGAATATTT AAAATTAGAG CTTGCTTAGT GTGCTGCTTC TGATCACTA
 5220
 TTGATCCCAC TTCGTGCTTT CATGTTAATT GGCCCAATTG GACTCTACAG TTGGAAGGT
 5280
 AAAACTTACT ATTTCAACTT GAGTCACGTA TGTATTCTTA TCATATACTT CTAAAGGT
 5340
 20 CTATTTTTT TCTTCTGATA GTCACCACAC CAAGCACTTC CAGCCACCCCT GCCACAGAC
 5400
 TCCTTTGTAA TCACTGTTGA AGGACATGAT GTTTTATGA CTTCCCGAAA TGAAAACCC
 5460
 ATCTTGTTTT TAAAACAAAC AAACCAACAA AAAGTAGTGT TTATGTAAGC ATTTGTTCC
 5520
 25 CTGACTCTAG GAACCCCTCT GTTTTATAT CAACTCTGTA CTGGCAAAAC ACAAAAACA
 5580
 AATGCCACCT TGCTAATTCC CTTCTAGCA AAGTAATACA GTTTAGCACA TGTTCAAGA
 5640
 30 AAAAATGGCT AAGAAATTTT GTTCCACTA ATTATTTCA AGACTGTGAT ATTTACACTC
 5700
 TGCTCTCAA ACGTTACATT TTATAAGACT ATTTTTAAC ATGTTGAACA TAAGCCCTA
 5760
 ATATATGTAT CCTTAAATTG TATTCAAAT ATTTAGGTC AGTCTTGCT ATCATTCCAC
 5820
 35 GAATAGAAAG TTTAACACT GGAAACTGCA AGTAAATATT TGCCCTTTA CCTGAATTI
 5880
 GGTAGCCCTC TCCCCAAGCT TACTTTCTGT TGCAGAAAGT GTAAAATTA TTACATAAA
 5940
 TTCTAATGAT GGTATCCGTG TGGCTTGCAT CTGATACAGC AGATAAAGAA GTTTTATGA
 6000
 AATGGACTCC TGTCCACTG AAAAGTAAT CTTAATGGCC TGTATCAACT ATCCTTGAC
 6060
 40 ACCATATTGA GCTTGGGAGG AAGGGAAAGT CCTGAATGAG GTTATAAAGT AAAAGAAAA
 6120
 ATTGCAAAA TGTCCCTTT TTTAAATGT TACATTTAG AAATATTAA AGTGTGTA
 6180
 45 CATTGTAGGA ATTACCCCAA TAGGACTGAT TATTCGCAT TGTAAAATAA GAAAAAGTTI
 6240
 TGTGCTGAAG TGTGACCAGG AAGTCTGAAA ATGAAGAGAG ACAGATGACA AAAGAAGATC
 6300
 CTTCTAATGG ACTAAGGAGG TGCTTCTTA AAGTCAGAAA GAGATACTCA GAAAGAGGTA
 6360
 50 CAGTTTTGG AAGGCACAGA GCCCCAACTT TTACGGAAGA AAAGATTCA TGAAAATAGI
 6420
 GATATTACAT TAAAAGAAAGT ACTCGTATCC TCTGCCACTT TATTCGACT TCCATTGCC
 6480
 TAGGAAAGAG CCTGTTGAA GGCGGGCCCA AGGAGTGCCG ACAGCAGTCT CCTCCCTCCA
 6540
 55 CCTCTTCCT CATTCTCTCC CCAGCTTGCT GAGCCCTTG CTCCCCTGGC GACTGCCTGC
 6600

1 ACAGTCAGCA AGGAATTGTC TCCCAGTGCA TTTTGCCTC CTGGCTGCCA ACTCTGGCTG
 6660
 2 CTAAGCGGC TGCCACCTGC TGCAGTCTAC ACAGCTTCGG GAAGAGGAAA GGAACCTCAG
 6720
 3 ACCTTCCAGA TCGCTTCCTC TCGCAACAAA CTATTTGTCG CAGGTAAGAA ATATCATTCC
 6780
 4 TCTTATTTG GAAAGTCAGC CATGGCAATT AGAGGTAAT AAGCTAGAAA GCAATTGAGA
 6840
 5 GGAATATAAA CCATCTAGCA TCACTACGAT GAGCAGTCAG TATCAACATA AGAAATATAA
 6900
 6 GCAAAGTCAG AGTAGAATT TTTCTTTA TCAGATATGG GAGAGTATCA CTTTAGAGGA
 6960
 7 GAGGTTCTCA AACTTTTGC TCTCATGTTC CCTTTACACT AAGCACATCA CATGTTAGCA
 7020
 8 TAAGTAACAT TTTAATTAA AAATAACTAT GTACTTTTT AACAAACAAAA AAAAGCATAA
 7080
 9 AGAGTGACAC TTTTTTATT TTACAAGTGT TTTAACTGGT TTAATAGAAG CCATATAGAT
 7140
 10 CTGCTGGATT CTCATCTGCT TTGCATTAG ACTACTGCAA TATTGCACAG AATGCAGCCT
 7200
 11 CTGGTAAACT CTGTTGTACA CTCATGAGAG AATGGGTGAA AAAGACAAAT TACGTCTTAG
 7260
 12 AATTATTAGA AATAGCTTC ACTTTAGGAA CTCCCTGAGA ATTGCTGCTT TAGAGTGGTA
 7320
 13 AGATAAATAA GCTTCTCTT AAACGGAATC TCAAGACAGA ATCAGTTACA TTAAAAGCAA
 7380
 14 ACAAAAAATT TGCCCATGGT TAGTCATCTT GTGAAATCTG CCACACCTT GGACTGGCT
 7440
 15 ACAATTGGAT AATATAGCAT TCCCCGAGAT AATTTCTCT CACAATTAAG GAAAGGGCTG
 7500
 16 AATAAAATATC TCTGTTGAA GTTGAATAAC AAAAATTAGG ACCCCCTAAA TTTTAGGGCT
 7560
 17 CCTGAAATTG GTCTTTTGC CTATATTCACTACG TTCTATTAAA TCTTCTTCA
 7620
 18 GGCCAGGTGC ACTAGCTCAT GCCTAGAATC TCAGGCAGGC CTGAGCCCAG GAATTGAGA
 7680
 19 CCAGCCAGGG CAACACAGTC TCTACAAAAA AATAAAAAT TACCTGGGTG TGTTGGTGCA
 7740
 20 TGCCTGTAGA ACTACTCAGG ATGCTGAGGA CTGCTTGAGC CCAGGATAGC CAAATCTGTG
 7800
 21 GTGAGTTCAAG CCACTAAACA GAGCGAGACT TTCTAAAAA AACAAACAAA AAAACAAACA
 7860
 22 AACCTCCTTC AAAATAACTT TTTATCTGCA ATGTTTCCT ATTGCCTGTG AGATTAATT
 7920
 23 TACTCTTTA CCTGATTTC CAAAGCCCTCC ATAATCTAAT CCGACTTTAC CTTGTGTTCA
 7980
 24 CTGAAAATA GCAGGACTGT TCCACTACAA TCCAAAAATC ACAGGTTGGG TGCAGTGGCT
 8040
 25 CACTCCTGTA ATCCCAACAC TTTGGAAGGC CAAGGCAGGT GGATTGCTTC AGCTCAGGAG
 8100
 26 TTCAAGACCA GCCTGGGCAA CATGGAAAA ACCCTGTCTC TCCAAAACAT ACAAAAATTA
 8160
 27 GCCAGATGTG GTAGTATGTG CCTGTAGTCC CAACTACTCA AAAGGCTAAG GCAAGAGGAT
 8220
 28 CACTTGAGCC CAGGAGGTCA AGGCTACAGT GAGCCATGTT TACTGTGTCA CTGCACTCCA
 8280
 29 GCCTGGGTGA TAGAGCAAGA CCATGTCTCA AAAAAAAA AAAGAAAAGA AAAGAAAAAA
 8340
 30 ACATCGCTCT ATTCAAGTTCA CCCCCACCAC AACATTGTT TGATTATCAC ATAAATGCTG
 8400
 31 GTCCATTGCC TTCTCTATCT ATTCAAAATCT TTAAGCATTC TTTGAGATTG AACTCAATTC

8460
 TCCTTTCAA ACTAGGCCAT TTAAACTACA TCAGTTCCAT TTTGATTTTC TTGCTTTGA
 8520
 5 TCTACAGACT CAAAAACAAA AACTAAAAA CTTATTTTT AAGTTTCTG CTACTCTCA
 8580
 TTCTCAACA CTCACATACA CGCATTATA ATAAGATGGC AGAATGTTCA AGGATAAAA
 8640
 GATTTATAGA ACTGAAAAGT TAGGTTTGA TCTTGTGCT GTCAAGATGA CTACCTACC
 8700
 10 GATCTCAGGT AATTAATTAT GTAGCATGCT CCCTCATTTC ATCCCATAACC TATTCAACA
 8760
 GATTGGAATT CCACAGCAAG GATAAACATA ATCATAGTTG CTTTCAAGT TCAAGGCAT
 8820
 TTAACCTTTA ATCTAGTAGT ATGTTGTTG TTCTTGTGTTGTTGAGAT GGAGCCCTG
 8880
 15 TGTGTCACCC AGGCTGGAGT GCAGTGGCAC GAACTCGGCT CACTGCAACC TCTGCCTCA
 8940
 GGGTTCAATC AGTTATTCTG CCTCAGTGTC CCAAGTAGCT GGGACTACAA GGCACATGC
 9000
 ACCATGCCCTG GCTAATTTT GTATTTTAG TAGAAACAGG GCTTCACCACAT GTGGCCAG
 9060
 20 CTGGTCTCGA ACTCCTGACC TCAAGTGATC CAGCCGCCTC GGCCTCCCAA AGTGTGGGG
 9120
 TTACAGGCAT AAGCCACCGT GCCCAGCCTA ATAGTATGTT TTTAAACTCT TAGTGGCTT
 9180
 25 ACAATGCTGG TTGTATAATA AATATGCCAT AAATATTTAC TGTCTTAGAA TTATGAAGA
 9240
 GTGGTTACTA GGCGTTTGC CACATATCAA TGTTCTCTC CTTACAGCTT TAATTAGAG
 9300
 CTAGAATTGC AGGTTGGTAG AGCTGGAACA GACCTTAAAG ATTGACTAGC CAACTTCCT
 9360
 30 GTCCAATGA GGGAACTGAG ACCCTTAAAAA TTAAGTACT TGCCCCAGAC AAAACTGGAA
 9420
 CTCATGTGTC CTAATTCCA TCATGAAATT CTACCATTCA CTAGCCTCTG GCTAGTTGTC
 9480
 AAAGTATTGC ATAACATAAT TTTTATGTC TTTTAAAGA ACAAAATTGTC ACTGCTTAC
 9540
 35 CCTGGGAGGG TCTTCTGAG GTGGTTTATA ACTCTTAAAAA AAAAAAAAGT CAGTAGTCTC
 9600
 AGAATTTAG ACGAAATAGT CAAAGCATT TTATCCAATG GATCTATAAT TTTCATAGA
 9660
 TAGAGTTAAA TCAAAGAACACGATGAGA AAGGAAGAGG AAAATTGAGG AGAGGAGGA
 9720
 40 TGGGGATGAG AACACACTAC TTGTAATCAG TCATAGATGT ACTGAGAACT AACAAAGAAG
 9780
 ATTGTAAGAA AATAAGAATG AAGAATTCAA AATCAACACA TGAAATAAAA AGAAACTAC
 9840
 AGGGAAAAAT GGAGAAGACA TTAGAAAAAT TATTCTATT TTAAAATTCT GTTTCAGGC
 9900
 45 TTCCCTCCTG TTCTCCTCC TTCTCATTGG TTTTCAGGTG GAGGGAAAGT TTAAGATGG
 9960
 AAAAATATAT ATATTCTACA CATCCCTTC TACGCTGTTG TCATGGCAAC AAGGTTTATC
 10020
 50 ATAGCAAATCT TTTATTCTATA CAACATTAT TGAGTTCTTA CTGTGTGGTA AGCTCTTCC
 10080
 AGGTGTTGAA AATTCAAGGG AAAAAGACA ACTCATTGTC TTAAAATCTA GATGAAAGC
 10140
 55 GAACAGACCT ATTTTAATC AAAGTAATCT CAATTTAGGG TAGTAAGAGC TATTTAAGAA
 10200
 GCATGAACAG GTGTGAAGGA GGTAGGACTC TGAGGAGAGA ATAGTTAGCT AGGAATGAAP
 10260

GAGCAGAGAA GTTTCTAG AGGAACATT AAAGCTGGGA GTTACGGGAT GAAAGATGAG
 10320
 5 GCAGGGTTG CAGGCAAAAA AAAAAAAAAG GCAGGGAAAG GGGAAAGTTCT GGCCTGGCAG
 10380
 AGAGAATAAC TGTGGCAACA ATGGAGGAGA GTCTGGAAGC AAGAAAACCA AGTAGAAGAG
 10440
 TATTAATAA GAAGATGCCA GGGGTAATGA GGGCTTGATT TAAAACAGTG CTGTTGGAGA
 10500
 10 TGGAGAGGAG ATACCAAATT CTGGAGACAT TTCTGAGTTA GAACCTACAG TATTATCAG
 10560
 ACAAGGGAAA GATTAGACAA AGGAGTTAAG AATGACTCCC AGGTTTCAGT TTGGGGCAGG
 10620
 TAACTAGGAC ATGTTTGAA AAGTAATGTA TTGGATCTCT TACCATTGGA ACTATGTATG
 10680
 15 TGGAGCCAAA TTAAAATTG TACATGTATA TAACTCTCCC CCCACCACCA GTAACACTT
 10740
 CCCTAACTCT CTACTTTGTA GCCAGACTTC CTAAAAGAAT AGTTTGTAGT CACTGTCTT
 10800
 ACTTTCCCC TCCCATTCTG TCCTAGATAT TTGTCACCT ACCATCTGCT GCCTCCACTT
 10860
 20 TACCCAAACT GTTCTACGGT TGCCCAAAAC TTCTAATTG CCAAATTCAA TGAACAAGTT
 10920
 TAAGCTTATA TGTAAATTAG GAGCTCTACA GTTTGATTTC GAGCAGCCCC TCCTGAAACC
 10980
 CTTTCTCTTT CGACTTCTGT GACACATCTC AGATTTACAA AACTGAACTA ATTATTTAC
 11040
 25 ACTTGAGCTG TATTTTCGTT CTTCTTCTT GATGAATGAG GTAACCACTC AACAAATTGC
 11100
 CCAAGCCAAA AACTACGAAG TCATCCTCAG TTCCCTCTTC TTCTGTTGA CCCACAACAG
 11160
 ATCAGCTGAG AAATCCCGCT GTTCTAGTATC TCTTGAATTTC ATTACCTTAA TTTATAGCCT
 11220
 30 CATCAACTCT TAATTGTTAA AATTACTTCA GTAGTTGTTG TCTGACCTCT GTCCAATCTT
 11280
 GTTCAATCAG GTCCATTCTT TTGTTCTTGG TGGTGGTGGT GGTGTTGACA GAGTTCGCT
 11340
 35 TTTGCTGCC AGGCTGAAGT GCAGTGGAGC ACTTCACTGC AACCACAGCC TCCTGGTTT
 11400
 AAGCAGTTCA CCCTCCCGAG TAGCTGGGAC TACAGGTATG TGCCACCACA CCCAGCTAAT
 11460
 TTTGTGTTT CAGTAGAGAC AGGGTTTCAC CATGTTGGTC AGGCTGGTCT CAAACTCCTG
 11520
 40 ACCTCAAGCA ATCCACCCAC CTCAGCCTCC CAAAGTGTG GGATTACAGG CATGAGCCAC
 11580
 TGCACACGGA CCAGATCCAT TGTTTATGTT GCTTCTAGAG TGAGTTTTA AAACACAAAT
 11640
 TTGACCATAT CTTCTCCAA TTTAAGTCAG TATTTTTTT TTCAGGAAAA AACAGTTCAA
 11700
 45 ACTCTTTAGT CTGCTTACAC AAGGCCTTG TAGTCTGACT CTTCTTCCA AGCTTCATC
 11760
 AAAGTATACT GCAAGTTACA TTTTATGTGA ATTGAATTAG GCAACGGTAT AAAAATTATA
 11820
 GTTTATATGG GCAAAATGGA AATAATGTTA ACTCTTCAA ATAGTTATC TAGAATGACA
 11880
 50 TAATTCAAA GCTGTCAGGT CAAATGAGTT ATAAACTGTT AACACTATTG CCACATGCAA
 11940
 GTGTCTCTTA TACTGGTAG AATTATCTGC TTCCATGTCA TTATTATGTA AATTAGACTT
 12000
 TAAATAACTC AGAAGTTCTT CAGACATACA GGTTATTATT GTGCTTTTA AACATAATT
 12060
 55 TAAATAATT TATATATGAT AATGTTATCC AAGTGCTAAG GGATGTATTG TTACTGCTGT

12120
 GCAAAAAAAA AAAAAAAA AACTCCAAAT AAATATGTTG AAACCAAGTT TATATGCAA
 12180
 5 AAAACAATAT TAAAAAGGCC AAAGTACCAAC CATAATAGGC TGTGTGGAGA CGGCAGGCT
 12240
 CAAAACACTA GTAATAATGC TGAGAAAGTT GAAAAAAGAA AGAAAGCAAC AATATGCTT
 12300
 10 GGTGTTGTA GGTTTATGTA CTCCAAGAAT ATCTCCTCTC AAACTTTAC GTTTTTCC
 12360
 AAGAAAAGTT AACTTGGCT GGGCGCAGTG GCTCTGCCT GTAGTCCCAG CCTTTGGGA
 12420
 GCAAGGCAGG GCAGATCACC TGAGGTCAGG AGTTGAGAC CAGCCTGACC AAAATGGA
 12480
 15 AAACCCGCC CCCTCACTAC TAAAAGAATA CAAAATTAGG CCGGGCACAG TGGCTTACC
 12540
 CTGTGATCCC AGCACTTGG GAGGCCGAAG CAGGAAGATC ACCTGAGGTC AGGAGTTCG
 12600
 GACCAGCCAT GGAGAAACCC GTCTCTACTA AAAATACAAA ATTAGCCGGG CGTGGTGGT
 12660
 20 CATGACTGTA ATCCCAGCTA CTCAGGAGGC TAAGGCAGAG AATCACTTGA ACCCAGGCA
 12720
 TGAGGTTGC AGTGAGCCGA GATCGTGCCA TTGCACTCCA GCCTGGCAA CAAGAGCGA
 12780
 ACTCTGTATC CAAAAAACAA AAGAAAAGAA AAGGTAACCT TGAACATATGT GAGATCTT
 12840
 25 GAAATGCATT CTTTCTGTAA AATGTGACTA CATTGCCTT ATTTATGGTA AAAATGTTG
 12900
 GGCTCAAAC AACCCATATT TTCTCGGTCT CCCCCGCTGCC TAGCCTTGT TCACATTGC
 12960
 TCTTCTGGT GGAAGCTCTT CCTCTGGCCT TGAAAATGCC TGCTTCTCTT TCAAGGTAG
 13020
 30 ACAGTCATCA CTTTCTGTGG TAACCTTCTC CAGCACCATC AAACAGAAAG AATGAATCT
 13080
 TTGTAAATTTC AGCTCTTACG TCATTCAATT CATTATTTG TAACTCTTTA TAGATTCTT
 13140
 TCTCCCCTACTA GACTCTGAGT CACTGGAGAG TAGGAGCCAA CTCTCATTCA TGTGTGGTT
 13200
 35 GGTCAAGCTAC TGGCCACATT CCTGATGCAT AGTTAATGCT CAAACCTTAA CTGGTGAAT
 13260
 AGCTCAAATA TTGTCTTCTT CTAAATCCAT TCACTCATTG ACTAACTATG TACTCAAAA
 13320
 AGTAAACACC AGTAATTAA TCCAATTCCCT GCCCATACTG CTTGGTACAT TTCAGGTGA
 13380
 40 TTGTTTGAT AAATATGTGT GTATTACATA ATATTAAAGT ATGTACAGAA GATCATGCT
 13440
 ATCATAATTTC ACAACTGATA ACTAATCAA CATAATGCT CTCAGGTTAA CAAATGTCT
 13500
 CCTTCTCAGT TAATGCAGTC ATTAACAAAC ACCTCTGAT GCTGATAATA GGGCCTTGT
 13560
 45 CAGCAATGAA GCCATAAAGG TGAATAAAGA ACATGCCCTC GTGGAGCTCA CAGCCTAGT
 13620
 ATTATTGTTG TGATTTTAA TATTAATGTT GGTTGGGTT TTGGTAAAAA ATGTTTAGAC
 13680
 50 TTATCTTAGT GATCTTTCA TCCTTGCTA TATTATTTT CTCTAAGAGT CTTCCTTATC
 13740
 CCCTCCTTTA AAAAAGTAGG TGATAATTCT AAATTGTAAA TTTAAATATT ATAAATAGC
 13800
 TATAAAATTT AATATTATA ATATTAAAT GTTGATAAAA TATTTAAATT TTATAATAT
 13860
 55 TAAATGTTA TTTAAATTCA TTTGTACATC AGTTTTTATT TTATTTAAAT GTGTTGGCC
 13920

GGCATGGTGG CTGACACCTA TAATCCCAGA ACTTTGAGAG GCCAAGTCAG GCAAACCATT
 13980
 TGAGCTCAGG AGTTTGAGAC CACCCCTGGGC AACGTGGTGA AACCTGTCT CTACCAAACA
 14040
 5 TATGAAAAC TATCTGGGTG TGGTGGCACG CATCTGTGGT CCCAGATGGG AGTCCCAGGC
 14100
 TAAGATGGGA GAATCGCTTG AACCCAGGTG AGAGGGGTGG GGTGGATGTT GCAGTGAGCT
 14160
 10 GAGATCGTGC CACTGCACTC CAACCTGGGT GACAGAGTGA GACTCCATCT CAAAAAAA
 14220
 AAATGTTATC TAAATAAGAT AAATTTAATA ACTGTTCGCA CTTAGATGAG CATAAGGAAC
 14280
 TAAACCTAGA TAAAACATAC AAATAAGGCC TGGGTACAGT GACTCATGCC TGTAATCTCA
 14340
 15 AGCACTTTGG GAGGCCAAAA TTATACAAAG TTAGTTGTAT AACACCAACT AACAACTATT
 14400
 TTGGGGTTAG CTTAATTCAAG ATTAATTTTT TTTAAACTGA GTTTTAAATT CCTGCTTACT
 14460
 CTACCATACA TGCTAGGCCT CATATTATGC TAGAAAAATT TTGAGCACAG ATTTATGAAT
 14520
 20 ACTCTCCTGC ATACCATTAA ATTTTAAAC AAATTTAAT GCAGTATATA TGTGCCTTT
 14580
 TACCAACACA TTAAATAATA AGATCTACTG TGAGGACTAA ATTTCTGTAA TTTCAAAGTA
 14640
 GTAATGAGTT TAAACCATGT CTCAAGATCT CTGCAATAAC TGTAGCACAA CAGAAAATAG
 14700
 25 GTATTCTAT TAATGACAGA GTCACAAAGTA CTACTAATAA TACTGTGGTT TGTTTCCTGC
 14760
 AACTAATCAT GGGAGGAATG CAAATTTCA GAGGTTGGTG AAAATACATG TGTATTTTT
 14820
 TCCCCATCCA AGTCACAGA TTTCTCACAC TGAGAACTCC TATTCCATAA CAAAATTCTG
 14880
 30 GAAGCCTGCA CACCGTATTG GAAGAAGGGC AGAAAGGAAA AGCAAATGGA AGGATTAAA
 14940
 TTTTTTCAA ATCCTGTATC CCTGATTTC ACAGCAAGAT TGTATTTATG TATTACTTGT
 15000
 GTTAAAATA TAGTATAATC GAGACTCCAG ATCAAAAATC ACCGCAGCTC AGGGAGAAAG
 15060
 35 AGGGCCACCA AATGCCAGAG CCCTTCAGCC TTCTCCCACC CTGCCTGTAC CCTCAGATGG
 15120
 AAGCACTTT TTATCATTGT TTCACCTTTA GCATTTGAC AATGAAGTCA CAAACCTTCA
 15180
 40 GCCTCTCACC CATAGGAACC CACTGGTTGT AAGAGAAGGA TGAAGCCAGT CCTTCCTAAA
 15240
 GGGCACGATT AGATGTGTTT ATGGCATCCT CAGGTGAAAC TATATTATA TTGACAATAT
 15300
 45 ATTTATATT CTCAAGGAAT ACTAGAATAA TGATTCAGTT CAGTACTAGG CCATTTATCT
 15360
 ACCCTTTATA ATATTGTTA ATGAGAAAAT GCTTTCTATC TTCCAAATAT CTGATGATT
 15420
 GTAAGAGAAC ACTTAAACAT GGGTATTCAAT AAGCTGAAAC TTCTGGCATT TATTGAATGT
 15480
 50 CAAGATTGTT CATCAGTATA CTAGGTGATT AACTGACCAC TGAACCTGAA GGTAGTATAA
 15540
 AGTAGTAGTA AAAGGTACAA TCATTGTCTC TTAACAGATG GCTCTTGCT TTCATTAGGA
 15600
 ATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA
 15651

Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala
 -35 -30 -25
 55 ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGGC TAATGCCATA

15702
 Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
 -20 -15 -10
 5 GAACAAATAC CAGGTTCAGA TAAATCTATT CAATTAGAAA AGATGTTGTG AGGTGAACTA
 15762
 TTAAGTGAATCTTGTCA CCAAATTCA CTGTAATATT AATGGCTCTT AAAAAAAATAG
 15822
 TGGACCTCTA GAAATTAAACC ACAACATGTC CAAGGTCTCA GCACCTTGTC ACACCACGTG
 15882
 10 TCCTGGCACT TTAATCAGCA GTAGCTCACT CTCCAGTTGG CAGTAAGTGC ACATCATGAA
 15942
 AATCCCAGTT TTCATGGAA AATCCCAGTT TTCATTGGAT TTCCATGGGA AAAATCCCAG
 16002
 TACAAAATG GGTGCATTCA GGAAATACAA TTTCCCAAAG CAAATTGGCA AATTATGTAA
 16062
 15 GAGATTCTCT AAATTTAGAG TTCCGTGAAT TACACCATT TATGTAAATA TGGTTGACAA
 16122
 GTAAAAAATTG ATTCTTTTT TTTTTTCTG TTGCCAGGC TGGAGTGCAG TGGCACAATC
 16182
 20 TCTGCTCACT GCAACCTCCA CCTCCTGGGT TCAAGCAATT CTCCTGCCTC AGCCTTCTGA
 16242
 GTAGCTGGGA CTACAGGTGC ATCCCGCCAT GCCTGGCTAA TTTTTGGGTA TTTTTACTAG
 16302
 AGACAGGGTT TTGGCATGTT GTCCAGGCTG GTCTTGGACT CCTGATCTCA GATGATCCTC
 16362
 25 CTGGCTCGGG CTCCCAAAGT GCTGGGATTA CAGGCATGAA CCACCACACA TGGCCTAAAA
 16422
 ATTGATTCTT ATGATTAATC TCCTGTGAAC AATTGGCTT CATTGAAAG TTTGCCTTCA
 16482
 TTTGAAACCT TCATTTAAAA GCCTGAGCAA CAAAGTGAGA CCCCATCTCT ACAAAAAAACT
 16542
 30 GCAAAATATC CTGTGGACAC CTCCTACCTT CTGTGGAGGC TGAAGCAGGA GGATCACTTG
 16602
 AGCCTAGGAA TTTGAGCCTG CAGTGAGCTA TGATCCACC CCTACACTCC AGCCTGCATG
 16662
 ACAGTAGACC CTGACACACA CACACAAAAA AAAACCTTCA TAAAAAATTA TTAGTTGACT
 16722
 35 TTTCTTAGGT GACTTCCGT TTAAGCAATA AATTAAAAG TAAAATCTCT AATTTTAGAA
 16782
 AATTATTTT TAGTTACATA TTGAAATTT TAAACCTAG GTTTAAGTTT TATGTCTAAA
 16842
 TTACCTGAGA ACACACTAAG TCTGATAAGC TTCATTTAT GGGCCTTTG GATGATTATA
 16902
 40 TAATATTCTG ATGAAAGCCA AGACAGACCC TAAACCATA AAAATAGGAG TTCGAGAAAG
 16962
 AGGAGTAGCA AAAGTAAAAG CTAGAATGAG ATTGAATTCT GAGTCGAAAT ACAAAATTT
 17022
 ACATATTCTG TTTCTCTCTT TTTCCCCCTC TTAG CT GAA GAT GAT G GTAAAGT
 45 17075
 Ala Glu Asp Asp Glu
 -10
 AGAAATGAAT TTATTTTCT TTGCAAACATA AGTATCTGCT TGAGACACAT CTATCTCACC
 17135
 50 ATTGTCAGCT GAGGAAAAAA AAAATGGTT CTCATGCTAC CAATCTGCCT TCAAAGAAAT
 17195
 GTGGACTCAG TAGCACAGCT TTGGAATGAA GATGATCATA AGAGATACAA AGAAGAACCT
 17255
 CTAGCAAAAG ATGCTTCTCT ATGCCTTAAA AAATTCTCCA GCTCTTAGAA TCTACAAAAT
 17315
 55 AGACTTTGCC TGTTTCATTG GTCCTAAGAT TAGCATGAAG CCATGGATTC TGTTGTAGGG
 17375

1 GGAGCGTTGC ATAGGAAAAA GGGATTGAAG CATTAGAATT GTCCAAAATC AGTAACACCT
 17435
 5 CCTCTCAGAA ATGCTTGGA AAGAACGCTG GAAGGTTCCG GGTTGGTGGT GGGGTGGGGC
 17495
 AGAAAATTCT GGAAGTAGAG GAGATAGGAA TGGGTGGGGC AAGAACGACCA CATTCAAGAGG
 17555
 10 CCAAAAGCTG AAAGAAACCA TGGCATTAT GATGAATTCA GGGTAATTCA GAATGGAAGT
 17615
 AGAGTAGGAG TAGGAGACTG GTGAGAGGAG CTAGAGTGAT AACACAGGGTG TAGAGCAAGA
 17675
 15 CGTTCTCTCA CCCCAAGATG TGAAATTGAG ACTTTATCTT GGAGATAATA GGGTTAATT
 17735
 AGCACAATAT GTATTAGCTA GGGTAAAGAT TAGTTGTTG TAACAAAGAC ATCCAAAGAT
 17795
 ACAGTAGCTG AATAAGATAG AGAATTTC TCTCAAAGAA AGTCTAAGTA GGCAGCTCAG
 17855
 20 AAGTAGTATG GCTGGAAGCA ACCTGATGAT ATTGGGACCC CCAACCTTCT TCAGTCTTGT
 17915
 ACCCATCATC CCCTAGTTGT TGATCTCACT CACATAGTTG AAAATCATCA TACTTCCTGG
 17975
 25 GTTCATATCC CAGTTATCAA GAAAGGGTCA AGAGAAGTCA GGCTCATTCC TTTCAAAGAC
 18035
 TCTAATTGGA AGTTAAACAC ATCAATCCCC CTCATATTCC ATTGACTAGA ATTTAATCAC
 18095
 ATGGCCACAC CAAGTGCAAG GAAATCTGGA AAATATAATC TTTATTCCAG GTAGCCATAT
 18155
 30 GACTCTTAA AATTCAAGAAA TAATATATTT TTAAAATATC ATTCTGGCTT TGGTATAAAG
 18215
 AATTGATGGT GTGGGGTGAG GAGGCCAAAA TTAAGGGTTG AGAGCCTATT ATTTAGTTA
 18275
 TTACAAGAAA TGATGGTGTC ATGAATTAAG GTAGACATAG GGGAGTGCTG ATGAGGAGCT
 18335
 35 GTGAATGGAT TTTAGAAACA CTTGAGAGAA TCAATAGGAC ATGATTTAGG GTTGGATTTG
 18395
 GAAAGGGAGAA GAAAGTAGAA AAGATGATGC CTACATTTT CACTTAGGCA ATTTGTACCA
 19455
 TTCAGTGAAA TAGGGAACAC AGGAGGAAGA GCAGGTTTG GTGTATACAA AGAGGAGGAT
 18515
 40 GGATGACGCA TTTCGTTTG GATCTGAGAT GTCTGTGGAA CGTCCTAGTG GAGATGTCCA
 18575
 CAAACTCTTC TACATGTGGT TCTGAGTTCA GGACACAGAT TTGGGCTGGA GATAGAGATA
 18635
 TTGTAGGCTT ATACATAGAA ATGGCATTG AATCTATAGA GATAAAAAGA CACATCAGAG
 18695
 45 GAAATGTGTA AAGTGAGAGA GGAAAAGCCA AGTACTGTGC TGGGGGGAAT ACCTACATTT
 18755
 AAAGGATGCA GTAGAAAGAA GCTAATAAAC AACAGAGAGC AGACTAACCA AAAGGGGAGA
 18815
 50 AGAAAAACCA AGAGAATTCC ACCGACTCCC AGGAGAGCAT TTCAAGATTG AGGGGATAGG
 18875
 TGTTGTGTTG AATTTGCAG CCTTGAGAAT CAAGGGCCAG AACACAGCTT TTAGATTTAG
 18935
 CAACAAGGAG TTTGGTGATC TCAGTGAAAG CAGCTTGATG GTGAAATGGA GGCAGAGGCA
 18995
 55 GATTGCAATG AGTGAACAG TGAATGGAA GTGAAGAAAT GATACAGATA ATTCTTGCTA
 19055
 AAAGCTTGGC TGTTAAAGG AGGAGAGAAA CAAGACTAGC TGCAAAGTGA GATTGGGTTG
 19115
 ATGGAGCAGT TTTAAATCTC AAAATAAAGA GCTTTGTGCT TTTTGATTA TGAAAATAAT
 19175
 GTGTTAATTG TAACTAATTG AGGCAATGAA AAAAGATAAT AATATGAAAG ATAAAATAT

19235
 AAAAACCA CAGAAATAAT GATAGCTACC ATTTGATAC AATATTCTA CACTCCTTC
 19295
 5 TATGTATATA TACAGACACA GAAATGCTTA TATTTTATT AAAAGGGATT GTACTATACC
 19355
 TAAGCTGCTT TTTCTAGTTA GTGATATATA TGGACATCTC TCCATGGCAA CGAGTAATTG
 19415
 CAGTTATATT AAGTCATGA TATTCACAA TAAGGGCATA TCTTTGCCCT TTTTATTTAA
 19475
 10 TCAATTCTTA ATTGGTGAAT GTTGTTC AGTTGTTGT TGTTATTAAC AATGTTCCCA
 19535
 TAAGCATTCC TGTACACCAA TGTCACACA TTTGCTGAT TTTTCTTCA GGATAAAACC
 19595
 CAGGAGGTAG AATTGCTGGG TTGATAGAAG AGAAAGGATG ATTGCCAAAT TAAAGCTTCA
 19655
 15 GTAGAGGGTA CATGCCGAGC ACAAAATGGGA TCAGCCCTAG ATACCAGAAA TGGCACTTTC
 19715
 TCATTTCCCC TTGGGACAAA AGGGAGAGAG GCAATAACTG TGCTGCCAGA GTTAAATTTG
 19775
 TACGTGGAGT AGCAGGAAAT CATTGCTGA AAATGAAAAC AGAGATGATG TTGTAGAGGT
 19835
 20 CCTGAAGAGA GCAAAGAAAA TTTGAAATTG CGGCTATCAG CTATGGAAGA GAGTGCTGAA
 19895
 CTGAAAACA AAAGAAGTAT TGACAATTGG TATGCTTGTATGGCACCGA TTTGAACGCT
 19955
 25 TGTGCCATTG TTCACCAGCA GCACTCAGCA GCCAAGTTG GAGTTTGTATGGCACCGA
 20015
 AAATAAGTTA GGGATTAAAT ATCCTGGCCA AATGGTAGAC AAAATGAACCT CTGAGATCCA
 20075
 GCTGCACAGG GAAGGAAGGG AAGACGGGAA GAGGTTAGAT AGGAAATACA AGAGTCAGGA
 20135
 30 GACTGGAAGA TGTTGTGATA TTTAAGAACAA CATAGAGTTG GAGTAAAGT GTAAGAAAAC
 20195
 TAGAAGGGTA AGAGACCGGT CAGAAAGTAG GCTATTGAA GTTAACACTT CAGAGGCAGA
 20255
 GTAGTTCTGA ATGGTAACAA GAAATTGAGT GTGCCTTGA GAGTAGTTA AAAAACAAATA
 20315
 35 GGCAACTTTA TTGTAGCTAC TTCTGGAACAA GAAGATTGTC ATTAATAGTT TTAGAAAAC
 20375
 AAAATATATA GCATACTTAT TTGTCAATTAA ACAAAAGAAAAC TATGTATTTT TAAATGAGAT
 20435
 TTAATGTTTA TTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA
 20486
 40 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu
 -5 1 5
 TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT
 20534
 Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile
 45 10 15 20
 GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT
 20582
 Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys
 25 30 35
 50 AGA G GT ATTTTTTTA ATTGCAAAC ATAGAAATGA CTAGCTACTT CTTCCCATTC
 20638
 Arg Asp
 40
 TGTTTTACTG CTTACATTGT TCCGTGCTAG TCCCAATCCT CAGATGAAAA GTCACAGGAG
 20698
 TGACAATAAT TTCACTTACA GGAAACTTTA TAAGGCATCC ACGTTTTTA GTTGGGGTAA
 20758

AAAATTGGAT ACAATAAGAC ATTGCTAGGG GTCATGCCTC TCTGAGCCTG CCTTTGAATC
 20818
 ACCAATCCCT TTATTGTGAT TGCATTAACT GTTTAAAACC TCTATAGTTG GATGCTTAAT
 20878
 5 CCCTGCTTGT TACAGCTGAA AATGCTGATA GTTTACCAGG TGTGGTGGCA TCTATCTGTA
 20938
 ATCCTAGCTA CTTGGGAGGC TCAAGCAGGA GGATTGCTTG AGGCCAGGAC TTTGAGGCTG
 20998
 TAGTACACTG TGATCGTACC TGTGAATAGC CACTGCACTC CAGCCTGGGT GATATACAGA
 21058
 10 CCTTGTCTCT AAAATTAAAAA AAAAAAAA AAAAAACCTT AGGAAAGGAA ATTGATCAAG
 21118
 TCTACTGTGC CTTCCAAAAC ATGAATTCCA AATATCAAAG TTAGGCTGAG TTGAAGCAGT
 21178
 15 GAATGTGCAT TCTTTAAAAA TACTGAATAC TTACCTTAAC ATATATTAA AATATTTAT
 21238
 TTAGCATTAA AAGTTAAAAA ACAATCTTT AGAATTCAA TCTTTAAAAT ACTCAAAAAA
 21298
 GTTGCAGCGT GTGTGTTGTA ATACACATTA AACTGTGGGG TTGTTGTTT GTTGAGATG
 21358
 20 CAGTTCACT CTGTCACCCA GGCTGAAGTG CAGTGCAGTG CAGTGGTGTG ATCTCGGCTC
 21418
 ACTACAAACCT CCACCTCCCA CGTTCAAGCG ATTCTCATGC CTCAGTCTCC CGAGTAGGTG
 21478
 GGATTACAGG CATGCACCAC TTACACCCGG CTAATTTTG TATTTTTAGT AGAGCTGGGG
 21538
 25 TTTCACCATG TTGGCCAGGC TGGTCTCAA CCCCTAACCT CAAAGTGTACT GCCTGCCTCA
 21598
 GCCTCCAAA CAAACAAACA ACCCCACAGT TTAATATGTG TTACAACACA CATGCTGCAA
 21658
 CTTTTATGAG TATTTAATG ATATAGATTA TAAAAGGTTG TTTTTAACCT TTAAATGCTG
 21718
 30 GGATTACAGG CATGAGCCAC TGTGCCAGGC CTGAACGTG TTTTTAAAAA TGTCTGACCA
 21778
 GCTGTACATA GTCTCCTGCA GACTGGCCAA GTCTCAAAGT GGGAACAGGT GTATTAAGGA
 21838
 35 CTATCCTTTG GTTAAATTTC CGCAAATGTT CCTGTGCAAG AATTCTTCTA ACTAGAGTTC
 21898
 TCATTTATTA TATTTATTTC AG AT AAT GCA CCC CGG ACC ATA TTT ATT ATA
 21949

Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile
 40 45

40 AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT
 21997
 45 Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile Ser
 50 55 60 65
 GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT
 22045
 50 Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile
 70 75 80
 TCC TTT AAG GTAAGACTG AGCCTTACTT TGTTTCAAT CATGTTAATA TAATCAATAT
 22103
 55 Ser Phe Lys
 AATTAGAAAT ATAACATTAT TTCTAATGTT AATATAAGTA ATGTAATTAG AAAACTCAAA
 22163
 TATCCTCAGA CCAACCTTT GTCTAGAACAA GAAATAACAA GAAGCAGAGA ACCATTAAAG
 22223
 TGAATACTTA CTAAAATTA TCAAAACTCTT TACCTATTGT GATAATGATG GTTTTCTGA
 22283
 GCCTGTCACA GGGGAAGAGG AGATACAACA CTTGTTTAT GACCTGCATC TCCTGAACAA
 22323

1 TCAGTCTTTA TACAAATAAT AATGTAGAAT ACATATGTGA GTTATACATT TAAGAATAAC
 22403
 5 ATGTGACTTT CCAGAATGAG TTCTGCTATG AAGAATGAAG CTAATTATCC TTCTATATTT
 22463
 CTACACCTTT GTAAATTATG ATAATATTTT AATCCCTAGT TGTTTGTG CTGATCCTTA
 22523
 GCCTAAGTCT TAGACACAAG CTTCAGCTTC CAGTTGATGT ATGTTATTT TAATGTTAAT
 22583
 10 CTAATTGAAT AAAAGTTATG AGATCAGCTG TAAAAGTAAT GCTATAATTA TCTTCAAGCC
 22643
 AGGTATAAAG TATTCTGGC CTCTACTTT TCTCTATTAT TCTCCATTAT TATTCTCTAT
 22703
 TATTCTCTC TATTCCCTCC ATTATTGTTA GATAAACCAC AATTAACAT AGCTACAGAC
 22763
 15 TGAGCCAGTA AGAGTAGCCA GGGATGCTTA CAAATTGGCA ATGCTTCAGA GGAGAATTCC
 22823
 ATGTCATGAA GACTCTTTT GAGTGGAGAT TTGCCAATAA ATATCCGCTT TCATGCCAC
 22883
 CCAGCTCCCCA CTGAAAGACA GTAGGATAT GACCTTAGTG AAGGTACCAA GGGGCAACTT
 22943
 20 GGTAGGGAGA AAAAAGCCAC TCTAAAATAT AATCCAAGTA AGAACAGTGC ATATGCAACA
 23003
 GATACAGCCC CCAGACAAAT CCCTCAGCTA TCTCCCTCCA ACCAGAGTGC CACCCCTTCA
 23063
 GGTGACAATT TGGAGTCCCC ATTCTAGACC TGACAGGCAG CTTAGTTATC AAAATAGCAT
 23123
 25 AAGAGGCCTG GGATGGAAGG GTAGGGTGG AAGGGTTAAG CATGCTGTTA CTGAACAACA
 23183
 TAATTAGAAG GGAAGGAGAT GGCCAAGCTC AAGCTATGTG GGATAGAGGA AAACTCAGCT
 23243
 GCAGAGGCAG ATTCAAGAAC TGGGATAAGT CCGAACCTAC AGGTGGATTC TTGTTGAGGG
 23303
 30 AGACTGGTGA AAATGTTAAG AAGATGGAAA TAATGCTTGG CACTTAGTAG GAACTGGCA
 23363
 AATCCATATT TGGGGGAGCC TGAAGTTTAT TCAATTGTA TGGCCCTTT AAATAAAAAG
 23423
 AATGTGGCTG GGC GTGGTGG CTCACACCTG TAATCCCAGC ACTTGGGAG GCCGAGGGGG
 23483
 35 GCGGATCACC TGAAGTCAGG AGTTCAAGAC CAGCCTGACC AACATGGAGA AACCCCATCT
 23543
 CTACTAAAAA TACAAAATTA GCTGGCGTG GTGGCATATG CCTGTAATCC CAGCTACTCG
 23603
 40 GGAGGCTGAG GCAGGAGAAT CTTTGAAACC CGGGAGGCAG AGGTTGCGAT GAGCCTAGAT
 23663
 CGTGCCATTG CACTCCAGCC TGGGCAACAA GAGCAAAACT CGGTCTAAA AAAAAAAA
 23723
 AAAAAAGTGAA ATTAACCAAA GGCATTAGCT TAATAATTAA ATACTGTTT TAAGTAGGGC
 23783
 45 GGGGGGTGGC TGGAAGAGAT CTGTGTAAT GAGGGAAATCT GACATTTAAG CTTCATCAGC
 23843
 ATCATAGCAA ATCTGCTTCT GGAAGGAACT CAATAAATAT TAGTTGGAGG GGGGGAGAGA
 23903
 GTGAGGGGTG GACTAGGACC AGTTTAGCC CTTGTCTTTA ATCCCTTTTC CTGCCACTAA
 23963
 50 TAAGGATCTT AGCAGTGGTT ATAAAAGTGG CCTAGGTTCT AGATAATAAG ATACAACAGG
 24023
 CCAGGCACAG TGGCTCATGC CTATAATCCC AGCACTTTGG GAGGGCAAGG CGAGTGTCTC
 24083
 ACTTGAGATC AGGAGTTCAA GACCAGCCTG GCCAGCATGG CGATACTCTG TCTCTACTAA
 24143
 55 AAAAAATACA AAAATTAGCC AGGCATGGTG GCATGCACCT GTAATCCCAG CTACTCGTGA

24203
 GCCTGAGGCA GAAGAACCGC TTGAAACCAAG GAGGTGAGG CTGCAGTGAG CTGAGATCGC
 24263
 5 ACCACTGCAC TCCAGCCTGG GCGACAGAAT GAGACTTGT CTCAAAAAAA GAAAAAGATA
 24323
 CAACAGGCTA CCCTTATGTG CTCACCTTTC ACTGTTGATT ACTAGCTATA AAGTCCTATA
 24383
 AAGTTCTTG GTCAAGAACC TTGACAACAC TAAGAGGGAT TTGCTTGAG AGGTTACTGT
 24443
 10 CAGAGTCTGT TTCATATATA TACATATACA TGTATATATG TATCTATATC CAGGCTTGGC
 24503
 CAGGGTTCCC TCAGACTTTC CAGTGCACCTT GGGAGATGTT AGGTCAATAT CAACTTCCC
 24563
 15 TGGATTTCAGA TTCAACCCCT TCTGATGTAA AAAAAAAA AAAAAAGAAA GAAATCCCTT
 24623
 TCCCCTTGGA GCACTCAAGT TTCACCAGGT GGGGCTTCC AAGTTGGGG TTCTCCAAGG
 24683
 TCATTGGGAT TGCTTCACA TCCATTTGCT ATGTACCTTC CCTATGATGG CTGGGAGTGG
 24743
 20 TCAACATCAA AACTAGGAAA GCTACTGCCA AAGGATGTCC TTACCTCTAT TCTGAAATGT
 24803
 GCAATAAGTG TGATTAAAGA GATTGCCTGT TCTACCTATC CACACTCTCG CTTTCAACTG
 24863
 TAACTTCTT TTTTCTTTT TTTCTTTTT TCTTTTTTT TGAAACGGAG TCTCGCTCTG
 24923
 25 TCGCCCAGGC TAGAGTGCAG TGGCACGATC TCAGCTCACT GCAAGCTCTG CCTCCCGGGT
 24983
 TCACGCCATT CTCCTGCCTC ACCCTCCAA GCAGCTGGGA CTACAGGCGC CTGCCACCAT
 25043
 GCCCAGCTAA TTTTTGTAT TTTTAGTAGA GACGGGGTTT CACCGTGTAA GCCAGGATGG
 25103
 30 TCTCGATCTC CTGAACCTGT GATCCGCCCG CCTCAGCCTC CCAAAGTGCT GGGATTACAG
 25163
 GCGTGAGCCA TCGCACCCGG CTCAACTGTA ACTTTCTATA CTGGTTCATC TTCCCTGTA
 25223
 ATGTTACTAG AGCTTTGAA GTTTGGCTA TGGATTATTT CTCATTTATA CATTAGATTT
 25283
 35 CAGATTAGTT CCAAATTGAT GCCCACAGCT TAGGGTCTCT TCCTAAATTG TATATTGTAG
 25343
 ACAGCTGCAG AAGTGGGTGC CAATAGGGGA ACTAGTTTAT ACTTCATCA ACTTAGGACC
 25403
 CACACTTGTGTT GATAAACGAA ACAGGTCAAG AGTTATGACT ACTGATTCCA CAACTGATTG
 25463
 40 AGAAGTTGGA GATAACCCCG TGACCTCTGC CATCCAGAGT CTTTCAGGCA TCTTTGAAGG
 25523
 ATGAAGAAAT GCTATTTAA TTTTGGAGGT TTCTCTATCA GTGCTTAGGA TCATGGGAAT
 25583
 45 CTGTGCTGCC ATGAGGCCAA ATTAAGTCC AAAACATCTA CTGGTTCCAG GATTAACATG
 25643
 GAAGAACCTT AGGTGGTGCC CACATGTTCT GATCCATCCT GCAAAATAGA CATGCTGCAC
 25703
 TAACAGGAAA AGTGCAGGCA GCACTACCAG TTGGATAACC TGCAAGATTA TAGTTCAAG
 25763
 50 TAATCTAACC ATTTCTCACA AGGCCCTATT CTGTGACTGA AACATACAAG AATCTGCATT
 25823
 TGGCCTCTA AGGCAGGGCC CAGCCAAGGA GACCATAATTC AGGACAGAAA TTCAAGACTA
 25883
 CTATGGAACG GGAGTGCTTG GCAGGGAAAGA CAGAGTCAAG GACTGCCAAC TGAGCCAATA
 25943
 55 CAGCAGGCTT ACACAGGAAC CCAGGGCCTA GCCCTACAAAC AATTATTGGG TCTATTCACT
 26003

GTAAGTTTA ATTCAGGCT CCACTGAAAG AGTAAGCTAA GATTCCCTGGC ACTTTCTGTC
 26063
 TCTCTCACAG TTGGCTCAGA AATGAGAACT GGTCAAGGCCA GGCATGGTGG CTTACACCTG
 26123
 5 GAATCCCAGC ACTTTGGGAG GCCGAAGTGG GAGGGTCACT TGAGGCCAGG AGTCAGGAC
 26183
 CAGCTTAGGC AACAAAGTGA GATACCCCCT GACCCCTTCT CTACAAAAAT AAATTTAAA
 26243
 10 AATTAGCCAA ATGTGGTGGT GTATACTTAC AGTCCCAGCT ACTCAGGAGG CTGAGGCAGG
 26303
 GGGATTGCTT GAGCCCAGGA ATTCAAGGCT GCAGTGAGCT ATGATTCAC CACTGCACTT
 26363
 CTGGCTGGC AACAGAGCGA GACCCCTGTCT CAAAGCAAAA AGAAAAAGAA ACTAGAACTA
 26423
 15 GCCTAAGTTT GTGGGAGGAG GTCATCATCG TCTTAGCCG TGAATGGTTA TTATAGAGGA
 26483
 CAGAAATTGA CATTAGCCA AAAAGCTTGT GGTCTTGCT GGAACCTCTAC TTAATCTTGA
 26543
 GCAAATGTGG ACACCACTCA ATGGGAGAGG AGAGAACTAA GCTGTTGAT GTATAGGGGA
 26603
 20 AAACTAGAGG CCTGGAAC TG AATATGCATC CCATGACAGG GAGAATAGGA GATTGGAGT
 26663
 TAAGAAGGAG AGGAGGTCAG TACTGCTGTT CAGAGATTT TTTTATGTAA CTCTTGAGAA
 26723
 GCAAAACTAC TTTTGTCTG TTTGTAATA TACTTCAAAA CAAACTTCAT ATATTCAAAT
 26783
 25 TGTTCATGTC CTGAAATAAT TAGGTAATGT TTTTTCTCT ATAG GAA ATG AAT CCT
 26839

Glu Met Asn Pro
85

CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA
 26887
 30 Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg
 90 95 100
 AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC
 26935
 35 Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr
 105 110 115 120
 GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC
 26983
 Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu
 125 130 135
 40 ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT
 27031
 Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr
 140 145 150
 GTT CAA AAC GAA GAC T AGCTATTAAA ATTCATGCC GGGCGCAGTG GCTCACGCCT
 27087
 45 Val Gln Asn Glu Asp
 155
 GTAATCCCAG CCCTTGAGGCG GGCTGAGGCG GGCAGATCAC CAGAGGTCAG GTGTTCAAGA
 27147
 CCAGCCTGAC CAACATGGTG AAACCTCATC TCTACTAAAA ATACAAAAAA TTAGCTGAGT
 27207
 50 GTAGTGACCC ATGCCCTCAA TCCCAGCTAC TCAAGAGGCT GAGGCAGGAG AATCACTTGC
 27267
 ACTCCGGAGG TGGAGGTTGT GGTGAGCCGA GATTGCACCA TTGCGCTCTA GCCTGGCAA
 27327
 CAACAGCAA ACTCCATCTC AAAAAATAAA ATAAATAAAAT AAACAAATAA AAAATTCTATA
 27387
 55 ATGTGAAC TG TCTGAATTT TATGTTAGA AAGATTATGA GATTATTAGT CTATAATTGT

27447
 AATGGTGAAA TAAAATAAAT ACCAGTCTTG AAAAACATCA TTAAGAAATG AATGAACCTT
 27507
 CACAAAAGCA AACAAACAGA CTTTCCCTTA TTTAAGTGAA TAAAATAAAA TAAAATAAAA
 5 27567
 TAATGTTAA AAAATTCAATA GTTTGAAAAC ATTCTACATT GTTAATTGGC ATATTAATTA
 27627
 TACTTAATAT AATTATTTT AAATCTTTG GGTTATTAGT CCTAATGACA AAAGATATTG
 27687
 10 ATATTTGAAC TTTCTAATTT TTAAGAATAT CGTTAAACCA TCAATATTT TATAAGGAGG
 27747
 CCACTTCACT TGACAAATTCTGAAATTCC TCCAAAGTCA GTATATTTT AAAATTCACT
 27807
 TTGATCCTGA ATCCAGCAAT ATATAAAAGG GATTATATAC TCTGGCCAAC TGACATTCA
 27867
 15 CCTAGGAATG CAAAGATGGT TTAATATCCT AAAATCAATT AACATAACAT ACTATATTAA
 27927
 TAAAGTATCA AAACAGTATT CTCATCTTTT TTTCTTTTT CACAATTCC TGGTTACACT
 27987
 ATCATCTCAA TAGATGCAGA AAAAGCATT GACAAAATCC AATTCTAAAT AAAAATTCTC
 28047
 20 AAACTTGAAA GAGAACATCA TAAAGGCATC TATGAAAAAC CTACAGCTAA TATCATACTT
 28107
 AACGATGAAA AACTGAATTA TTTTACCTA AGATCAAGAA TAATGCAAGC ATGTCAGCTC
 28157
 TTGCAACTTC TATTCAACAT TGTACTGGAG GTTCTAGCCA GAGCAACCAC ACAATAAATA
 28227
 25 AAAATAAAAG GCACCCAGAT TAGAAAGGAA GTCTTTATTT GCAGACAACA TGGTTCTTTA
 28287
 TGCAGAAAAC CGTCAGGAAT ACACACACAT GTTAGAACTA ATAAGTTCAAGC CAAGGTTGCA
 28347
 GCTTGCAATA TCAATATGCA AAAATACATT GAAGGCTGGG CTCAGTGGAG ATGGCATGTA
 30 28407
 CCTTTCGTCC CAGCTACTTG GGAGGCTGAG GTAGGAGGAT CACTTGAGGT GAGGAGTTG
 28467
 AGGCTATAGT GCAATGTGAT CTTGCCTGTG AATAGCCACT GCACTCGAGC CTAGGCAACA
 28527
 35 AAGTGAGACC CCGTCTCCAA AAAAAAAAT GGTATATTGG TATTTCTGTA TATGAACAAT
 28587
 GAATGATCTG AAAACAAGAA AATTCCATTC ACGATGGTAT TAAAAAAATA AAATACAAAT
 28647
 AAATTAGCA AAATAATTAT AAAACTTGTA CATCGAAAAT TTCAAAGCAC TCTGAGGGAA
 28707
 40 ATTAAAGATG ATCTAAATAA TTGGAGAGAC ACTCTATGAT CACTGATTGG AAAATTCA
 28767
 CAATATTGTT AAGATAACAA TTGTCCCCAA ATTGATGCAT GCATTCAATT TAGTCTTCAT
 28827
 CAAAATTCCA GCAGGGTTTT TGCAGAAATT GACAAGCTGT ACCCAAAATG TATATGGAAA
 28887
 45 TGAAAAGACC CAGAAGAGCA AATAATTGTT TAAAAACAAA GTTGGAAAAC TTTTACTTCC
 28947
 TAATTTAAA ACTTACTATA AACCTAAAGT TATCAAGACC ATTTAGT
 28994

50 (16) INFORMATION FOR SEQ ID NO:15:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH:10 amino acids
 (B) TYPE:amino acid
 (D) TOPOLOGY:linear
 55 (ii) MOLECULE TYPE:peptide

(v) FRAGMENT TYPE:N-terminal fragment
 (xi) SEQUENCE DESCRIPTION:SEQ ID NO:15:

5 Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser
 1 5 10

Claims

10

1. A genomic DNA, which encodes a polypeptide capable of inducing the production of interferon- γ by immunocompetent cells; said polypeptide possessing an amino acid sequence given in SEQ ID NO:1 (where the symbol "Xaa" means "isoleucine" or "threonine") or one of functional equivalents thereof;

15

SEQ ID NO: 1:

20 Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn
 1 5 10 15
 Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp
 20 25 30
 Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg Thr Ile Phe Ile
 35 40 45
 Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
 50 55 60
 Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys Glu Asn Lys Ile
 65 70 75 80
 Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys
 85 90 95
 Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys
 100 105 110
 Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu
 115 120 125
 Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu
 130 135 140
 35 Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp
 145 150 155.

40

2. The genomic DNA of claim 1, which comprises two or more exons; each of the exons possessing a part of nucleotide sequence given in SEQ ID NO:2;

SEQ ID NO: 2:

45 GCCTGGACAG TCAGCAAGGA ATTGTCTCCC AGTGCATTTT GCCCTCCTGG CTGCCAACTC 60
 TGGCTGCTAA AGCGGCTGCC ACCTGCTGCA GTCTACACAG CTTCGGGAAAG AGGAAAGGAA 120
 CCTCAGACCT TCCAGATCGC TTCCCTCTCGC AACAAACTAT TTGTCGCAGG AATAAAG 177
 ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA ATG 225

50

55

	Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala Met	
	-35 -30 -25	
5	AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA GAT GAT GAA AAC	273
	Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu Asp Asp Glu Asn	
	-20 -15 -10 -5	
	CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA GTC ATA	321
	Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile	
	1 5 10	
10	AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT CGG CCT	369
	Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro	
	15 20 25	
	CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA GAT AAT GCA CCC CGG	417
	Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg	
	30 35 40	
15	ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG	465
	Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met	
	45 50 55 60	
	GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA AYT CTC TCC TGT	513
	Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys	
	65 70 75	
20	GAG AAC AAA ATT ATT TCC TTT AAG GAA ATG AAT CCT CCT GAT AAC ATC	561
	Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile	
	80 85 90	
	AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA AGT GTC CCA GGA	609
	Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly	
	95 100 105	
25	CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC GAA GGA TAC TTT	657
	His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe	
	110 115 120	
	CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC ATT TTG AAA AAA	705
	Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys	
	125 130 135 140	
30	GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT GTT CAA AAC GAA	753
	Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu	
	145 150 155	
	GAC TAGCTATTAA AATTTCATGC CGGGCGCAGT GGCTCACGCC TGTAATCCCA	806
	Asp	
35	GCCCTTGGG AGGCTGAGGC GGGCAGATCA CCAGAGGTCA GGTGTTCAAG ACCAGCCTGA	866
	CCAACATGGT GAAACCTCAT CTCTACTAAA AATACTAAA ATTAGCTGAG TGTAGTGACG	926
	CATGCCCTCA ATCCCAGCTA CTCAAGAGGC TGAGGCAGGA GAATCACTTG CACTCCGGAG	986
	GTAGAGGTTG TGGTGAGCCG AGATTGCACC ATTGCCTCT AGCCTGGCA ACAACAGCAA	1046
	AACTCCATCT CAAAAAATAA AATAAATAA TAAACAAATA AAAAATTCTAT AATGTGAAAA	1106
	AAAAAAA AAAA	1120.

40 3. The genomic DNA of claim 1, which comprises two exons with respective nucleotide sequences given in SEQ ID NOs:3 and 4;

45 SEQ ID NO: 3:

AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA	47
Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser	

50

55

-5 1 5 10 95
 GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT
 Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn
 5 15 20 25
 CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA G 135
 Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp
 30 35 40

10 SEQ ID NO: 4:

AT AAT GCA CCC CGG ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC 47
 Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser
 40 45 50 55
 15 CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT 95
 Gln Pro Arg Gly Met Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile
 60 65 70
 TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT TCC TTT AAG 134
 Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile Ser Phe Lys
 80 85.

20

4. The genomic DNA of claim 1, which comprises two exons with respective nucleotide sequences given in SEQ ID NOs:5 and 6;

25

SEQ ID NO: 5:

GAATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG 50
 Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val
 -35 -30 -25
 30 GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G 87
 Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
 -20 -15 -10

35

SEQ ID NO: 6:

CT GAA GAT GAT G 12
 Ala Glu Asp Asp Glu
 -10.

40

5. The genomic DNA of claim 3, which comprises additional two exons with respective nucleotide sequences given in SEQ ID NOs:5 and 6.

6. The genomic DNA of claim 1, which comprises an exon with a part of a nucleotide sequence given in SEQ ID NO:7;

45

50

55

SEQ ID NO: 7:

5	GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile 85 90 95 100	48
10	TTC TTT CAG AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu 105 110 115	96
15	TCT TCA TCA TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp 120 125 130	144
20	CTT TTT AAA CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser 135 140 145	192
25	ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAAATTCA TGCCGGCGC Ile Met Phe Thr Val Gln Asn Glu Asp 150 155	246
30	AGTGGCTCAC GCCTGTAATC CCAGCCCTTT GGGAGGCTGA CGCGGGCAGA TCACCCAGAGG TCAGGTGTTA AAGACCAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA AAAATTAGCT GAGTGTAGTG ACCCATGCC TCAATCCCAG CTACTCAAGA GGCTGAGGCA GGAGAACATCAC TTGCACTCCG GAGGTGGAGG TTGTGGTAG CCGAGATTGC ACCATTCGCG TCTAGCCTGG GCAACAAACAG CAAAACCTCA TCTCAAAAAA TAAAATAAAT AAATAAACAA ATAAAAAAATT CATAATGTGA ACTGCTGAA TTTTTATGTT TAGAAAGATT ATGAGATTAT TAGTCTATAA TTGTAATGGT GAAATAAAAT AAATACCACT CTTGAAAAAC ATCATTAAGA AATGAATGAA CTTTCACAAA AGCAACACAA CAGACTTTCC CTTATTTAAG TGAATAAAAT AAAATAAAAT AAAATAATGT TTAAAAAATT CATAGTTGA AAACATTCTA CATTGTTAAT TGGCATATTA ATTATACCTA ATATAATTAT TTTTAAATCT TTTGGTTAT TAGTCTTAAT GACAAAAGAT ATTGATATTT GAACCTTCTA ATTTTTAAGA ATATCGTTAA ACCATCAATA TTTTTATAAG GAGGCCACTT CACTGACAA ATTTCTGAAT TTCCCTCCAA GTCAGTATAT TTTTAAAATT CAGTTTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC CAACTGACAT TCATCCTAGG AATCAGAAAGA TGGTTAATA TCCTAAAATC AATTAAACATA ACATACTATA TTAATAAAAGT ATCAAAACAG TATTCTCATC TTTTTCTT TTTTCACAAT TCCTGGTTA CACTATCATC TCAATAGATG CAGAAAAGC ATTTGACAAA ATCCAATTCA TAATAAAAAT TCTCAAACTT GAAAGAGAAC ATCATAAAAGG CATCTATGAA AAACCTACAG CTAATATCAT ACTTAACGAT GAAAAACTGA ATTATTTAC CCTAAGATCA AGAATAATGC AAGCATGTCA CCTCTTGCCTA CTTCTATTCA ACATTGTACT GGAGGTTCTA GCCAGAGCAA CCATACAATA AATAAAAATA AAAGGCACCC AGATTAGAA GGAAGTCTT ATTTGCAGAC 35 30 35 40 45	306 366 426 486 546 606 666 726 786 846 906 966 1026 1086 1146 1206 1266 1326 1386 1446 1506 1566 1626 1686 1746 1806 1866 1926 1986 2046 2106 2166 2167.

7. The genomic DNA of claim 3, which comprises additional one exon with a part of a nucleotide sequence given in SEQ ID NO:7.

8. The genomic DNA of claim 5, which comprises additional one exon with a part of a nucleotide sequence given in SEQ ID NO:7.

50 55 9. The genomic DNA of claim 1, which comprises two introns with respective nucleotide sequences given in SEQ ID NOs:8 and 9;

SEQ ID NO: 8:

5	GTATTTTTT TAATTCGCAA ACATAGAAAT GACTAGCTAC TTCTTCCCAT TCTGTTTAC 60
	TGCTTACATT GTTCCGTGCT AGTCCAATC CTCAGATGAA AAGTCACAGG AGTGACAATA 120
	ATTTCACTTA CAGGAAACTT TATAAGGCAT CCACGTTTT TAGTGGGGT AAAAAATGG 180
	ATACAATAAG ACATTGCTAG GGGTCATGCC TCTCTGAGCC TGCCTTGAA TCACCAATCC 240
	CTTTATTGTG ATTGCATTAA CTGTTAAAA CCTCTATAGT TGGATGCTTA ATCCCTGCTT 300
	GTACAGCTG AAAATGUTGA TAGTTTACCA GGTGTGGTGG CATCTATCTG TAATCCTAGC 360
10	TACTTGGGAG GCTCAAGCAG GAGGATTGCT TGAGGCCAGG ACTTTGAGGC TGTAGTACAC 420
	TGTGATCGTA CCTGTGAATA GCCACTGCAC TCCAGCCTGG GTGATATACA GACCTTGTCT 480
	CTAAAATTA AAAA AAAAAA AAAAAAACC TTAGGAAAGG AAATTGATCA AGTCTACTGT 540
	GCCTTCCAAA ACATGAATTC CAAATATCAA AGTTAGGCTG AGTTGAAGCA GTGAATGTCC 600
	ATTCTTAAA AATACTGAAT ACTTACCTTA ACATATATT TAAATATTTT ATTTAGCATT 660
	TAAAAGTTAA AAACAATCTT TTAGAATTCA TATCTTAAA ATACTCAAAA AAGTTGCAGC 720
15	GTGTGTGTTG TAATACACAT TAAACTGTGG GTGTTGTTGT TTGTTGAGA TGCA GTTCA 780
	CTCTGTCACC CAGGCTGAAG TGCAGTGCAG TGCAGTGGT TGATCTCGGC TCACTACAAC 840
	CTCCACCTCC CACGTTCAAG CGATTCTCAT GCCTCAGTCT CCCGAGTAGG TGGGATTACA 900
	GGCATGCACC ACTTACACCC GGCTAATTTT TGTTTTTA GTAGAGCTGG GGTTTCACCA 960
	TGTTGGCCAG GCTGGTCTCA AACCCCTAAC CTCAAGTGAT CTGCCTGCCT CAGCCTCCCA 1020
20	AACAAACAAA CAACCCACA GTTTAATATG TGTTACAACA CACATGCTGC AACTTTATG 1080
	AGTATTTAA TGATATAGAT TATAAAAGGT TGTTTTAAC TTTTAAATGC TGGGATTACA 1140
	GGCATGAGCC ACTGTGCCAG GCCTCAACTG TGTTTTAAA AATGCTGAC CAGCTGTACA 1200
	TAGTCTCCTG CAGACTGGCC AAGTCTCAAA GTGGGAACAG GTGTATTAAG GACTATCCTT 1260
	TGGTTAAATT TCCGCAAATG TTCCGTGCA AGAATTCTTC TAACTAGAGT TCTCATTAT 1320
	TATATTTATT TCAG 1334

25

SEQ ID NO: 9:

30	GTAAGACTGA GCCTTACTTT GTTTCAATC ATGTTAATAT AATCAATATA ATTAGAAATA 60
	TAACATTATT TCTAATGTTA ATATAAGTAA TGTAATTAGA AAACCTCAAAT ATCCTCAGAC 120
	CAACCTTTG TCTAGAACAG AAATAACAAG AAGCAGAGAA CCATTAAGT GAATACTTAC 180
	TAAAAATTAT CAAACTCTTT ACCTATTGTG ATAATGATGG TTTTCTGAG CCTGTCACAG 240
	GGGAAGAGGA GATACAACAC TTGTTTTATG ACCTGCATCT CCTGAACAAT CAGTCTTAT 300
	ACAAATAATA ATGTTAGATA CATATGTGAG TTATACATT AAGAATAACA TGTGACTTTC 360
	CAGAATGAGT TCTGCTATGA AGAATGAAGC TAATTATCCT TCTATATTTC TACACCTTG 420
35	TAAATTATGA TAATATTTA ATCCCTAGTT GTTTGTTGC TGATCCTTAG CCTAAGTCTT 480

40

45

50

55

5.	AGACACAAAGC TTCAGCTTCC AGTTGATGTA TGTTATTAA AATGTTAAC TAAATTGAATA 540 AAACTTATGA GATCAGCTGT AAAAGTAATG CTATAATTAT CTTCAAGCCA GCTATAAAAGT 600 ATTTCTGGCC TCTACTTTT CTCTATTATT CTCCATTATT ATTCTCTATT ATTTTTCTCT 660 ATTTCTCCA TTATTGTTAG ATAAACCACA ATTAACTATA GCTACAGACT GAGCCAGTAA 720 GAGTAGCCAG GGATGCTTAC AAATTGGCAA TGCTTCAGAG GAGAATTCCA TGTCAATGAAG 780 ACTCTTTTG AGTGGAGAT TGCCAATAAA TATCCGCTT CATGCCACC CAGTCCCCAC 840 TGAAAGACAG TTAGGATATG ACCTTAGTGA AGGTACCAAG GGGCAACTTG GTAGGGAGAA 900 AAAAGCCACT CTAAAATATA ATCCAAGTAA GAACAGTGCA TATGCAACAG ATACAGCCCC 960 CAGACAAATC CCTCAGCTAT CTCCCTCCA CCAGAGTGCC ACCCCCTTCAG GTGACAATT 1020 CGACTCCCCA TTCTAGACCT GACAGGCACC TTAGTATCA AAATAGCATA AGAGGCCTGG 1080 GATGGAAGGG TAGGGTGGAA AGGGTTAACG ATGCTGTTAC TGAACAAACAT ATTAGAAGG 1140 GAAGGAGATG GCCAAGCTCA AGCTATGTGG GATAGAGGAA AACTCAGCTG CAGAGGCAGA 1200 TTCAGAAACT GGGATAAGTC CGAACCTACA GGTGGATTCT TGTTGAGGGA GACTGGTGA 1260 AATGTTAAGA AGATGGAAT AATGCTTGGC ACTTAGTACG AACTGGCAA ATCCATATTT 1320 GGGGGAGCCT GAAGTTTATT CAATTGAT GCCCTTTTA AATAAAAAGA ATGTGGCTGG 1380 GCGTGGTGGC TCACACCTGT AATCCACCA CTTTGGGAGG CCGAGGGGG CGGATCACCT 1440 GAAGTCAGGA GTTCAAGACC AGCCTGACCA ACATGGAGAA ACCCCATCTC TACTAAAAT 1500 ACAAAATTAC CTGGGGCTGG TGGCATATGC CTGTAATCCC AGCTACTCGG GAGGCTGAGG 1560 CAGGAGAACAT TTTGAAACCC CGGAGGCAGA GTTGCAGATG AGCCTAGATC GTGCCATTGC 1620 ACTCCAGCCT GGGCAACAAG AGCAAAACTC GGTCTCAAAA AAAAAAAA AAAAGTGA 1680 20 TTAACCAAAG GCATTAGCTT AATAATTAA TACTGTTTT AAGTAGGGCG GGGGGTGGCT 1740 GGAAGAGATC TGTTGAAATG AGGGAATCTG ACATTTAACG TTCATCAGCA TCATAGCAA 1800 TCTGCTCTG GAAGGAACTC AATAAATATT AGTTGGAGGG GGGGAGAGAG TGAGGGTGG 1860 ACTAGGACCA GTTTTAGCCC TTGTCCTTAA TCCCTTTTCC TGCCACTAAT AAGGATCTA 1920 GCAGTGGTTA TAAAAGTGGC CTAGGTTCTA GATAATAAGA TACAACAGGC CAGGCACAGT 1980 GCCTCATGCC TATAATCCC GCACTTCCG AGGGCAAGGC CAGTCTCTCA CTTGAGATCA 2040 25 CGACTTCAG ACCAGCCTGG CCAGCATGGC GATACTCTGT CTCTACTAAA AAAAAATACAA 2100 AAATTAGCCA GGCATGGTGG CATGCACCTG TAATCCCACG TACTCGTAG CCGAGGGCAG 2160 AAGAATCGCT TGAACACCAGG AGGTGTTAGGC TGCAGTGAGC TGAGATCGCA CCACTGCACT 2220 CCAGCCTGGG CGACAGAATG AGACTTGTG TCAAAAAAAG AAAAGATAC AACAGGCTAC 2280 CCTTATGTGC TCACCTTCTA CTGTTGATTA CTAGCTATAA AGTCTTATAA AGTTCTTGG 2340 TCAAGAACCT TCACAACACT AAGAGGGATT TGCTTGAGA GTTACTGTG AGAGTCTGTT 2400 30 TCATATATAT ACATATACAT GTATATATGT ATCTATATCC AGGCTGGCC AGGGTTCCCT 2460 CAGACTTCC AGTGCACTTG GGAGATGTTA GGTCAATATC AACTTCCCT GGATTCAAGT 2520 TCAACCCCTT CTGATGTAAA AAAAAAAA AAAAGAAAG AAATCCCTT CCCCTGGAG 2580 CACTCAAGTT TCACCCAGGTG GGGCTTCCA AGTTGGGGGT TCTCCAAGGT CATTGGGATT 2640 GCTTTCACAT CCATTTGCTA TGTACCTTCC CTATGATGGC TGGGAGTGGT CAACATCAA 2700 35 ACTAGGAAAG CTACTGCCA AGGATGTCCT TACCTCTATT CTGAAATGTG CAATAAGTGT 2760 GATTAAGAG ATTCGCTGT CTACCTATCC ACACTCTCC TTTCAACTGT AACTTTCTT 2820 TTTTCTTTT TTCTTTTTT CTTTTTTT GAAACGGAGT CTCGCTCTGT CGCCCGAGGT 2880 AGAGTGCAGT GGCACGATCT CAGCTCACTG CAAGCTCTGC CTCCCCGGTT CACGCCATTC 2940 TCCTGCCCTCA CCCCTCCAAG CAGCTGGGAC TACAGGCAGC TGCCACCATG CCCAGCTAAT 3000 TTTTGTATT TTTAGTAGAG ACGGGTTTC ACCGTGTTAG CCAGGATGGT CTCGATCTCC 3060 40 TGAACCTCTG ATCCGCCCGC CTCAGCCTCC CAAAGTCCTG GGATTACAGG CGTCAGCCAT 3120 CGCACCCGGC TCAACTGTAA CTTTCATAC TGGTTCATCT TCCCTCTAA TGTTACTAGA 3180 GCTTTGAAG TTTGGCTAT GGATTATTC TCATTTATAC ATTAGATTTC AGATTAGTTC 3240 CAAATTGATG CCCACAGCTT AGGGTCTCTT CCTAAATTGT ATATTGTAGA CAGCTGCAGA 3300 ACTGGGTGCC AATAGGGGAA CTAGTTATA CTTTCATCAA CTTAGGACCC ACACCTGTG 3360 ATAAAGAACAA AAGGTCAAGA GTTATGACTA CTGATTCCAC AACTGATTGA GAAGTTGGAG 3420 45 ATAACCCCGT GACCTCTGCC ATCCAGAGTC TTTCAGGCAT CTTTGAAGGA TGAAGAAATG 3480 CTATTTAAT TTGGAGGTT TCTCTATCG TGCTTAGGAT CATGGAATC TGTGCTGCCA 3540 TGAGGCCAAA ATTAAGTCCA AAACATCTAC TGTTCCAGG ATTAACATGG AAGAACCTTA 3600 GGTGGTCCCC ACATGTTCTG ATCCATCTG CAAAATAGAC ATGCTGCAGT AACAGGAAA 3660 GTGCAGGCG CACTACCAGT TGGATAACCT GCAAGATTAT AGTTCAAGT AATCTAACCA 3720 50 TTTCTCACAA GGCCTATTCT TGTGACTGAA ACATACAAGA ATCTGCATT GGCCCTTCTAA 3780 GGCAGGGCCC AGCCAAGGAG ACCATATTCA GGACAGAAAT TCAAGACTAC TATGGAACCTG 3840 GAGTGCTTGG CAGGGAAAGAC AGAGTCAAGG ACTGCCAACT GAGCCAATAC AGCAGGCTA 3900
----	--

5 CACAGGAACC CAGGGCCTAG CCCTACAACA ATTATTGGGT CTATTCACTG TAACTTTAA 3960
 TTTCAGGCTC CACTGAAAGA GTAAGCTAAG ATTCCTGGCA CTTTCTGTCT CTCTCACAGT 4020
 TGGCTCAGAA ATGAGAACTG GTCAGGCCAG GCATGGTGGC TTACACCTGG AATCCCAGCA 4080
 CTTTGGGAGG CCGAAGTGGG AGGGTCACTT GAGGCCAGGA GTTCAGGACC AGCTTAGGCA 4140
 10 ACAAAGTGAG ATACCCCCCTG ACCCCTTCTC TACAAAATA AATTTAAAAA ATTAGCAAA 4200
 TGTGGTGGTG TATACTTACA GTCCCAGCTA CTCAGGAGGC TGAGGCAGGG GGATTGCTTG 4260
 AGCCCAGGAA TTCAAGGCTG CAGTGAGCTA TGATTTCAAC ACTGCACTTC TGGCTGGCA 4320
 ACAGACCGAG ACCCTGTCTC AAAGCAAAA GAAAAAGAAA CTAGAACTAG CCTAAGTTG 4380
 TGGGAGGAGG TCATCATCGT CTTTAGCCGT GAATGGTTAT TATAGAGGAC AGAAATTGAC 4440
 15 ATTAGCCAA AAAGCTGTG GTCTTGCTG GAACTCTACT TAATCTTGAG CAAATGTGGA 4500
 CACCACTCAA TGGGAGAGGA GAGAAGTAAG CTGTTTGATG TATAGGGAA AACTAGAGGC 4560
 CTGGAACCTGA ATATGCATCC CATGACAGGG AGAATAGGAG ATTCCGGAGTT AAGAAGGAGA 4620
 GGAGGTCAGT ACTGCTGTT AGAGATTTT TTTATGTAAC TCTTGAGAAG CAAAACACT 4680
 TTTGTTCTGT TTGGTAATAT ACTTCAAAAC AAACCTTCATA TATTCAAATT GTTCATGTCC 4740
 15 TGAAATAATT AGGTAATGTT TTTTCTCTA TAG 4773.

20 10. The genomic DNA of claim 1, which comprises three introns with respective nucleotide sequences of SEQ ID NOs: 10 to 12 as introns;

20 SEQ ID NO: 10:

 25 GTAAGAAATA TCATTCCCTCT TTATTTGGAA AGTCAGCCAT GCCAATTAGA GGTAAATAAG 60
 CTAGAAAGCA ATTGAGAGGA ATATAAACCA TCTAGCATCA CTACGATGAG CAGTCAGTAT 120
 CAACATAAGA AATATAAGCA AAGTCAGAGT AGAATTTTT TCTTTTATCA GATATGGGAG 180
 AGTATCACTT TAGAGGAGAG GTTCTCAAAC TTTTGCTCT CATTTTCCCT TTACACTAAG 240
 CACATCACAT GTTAGCATAA GTAACATTT TAATTAAGAA TAACTATGTA CTTTTTTAAC 300
 AACAAAAAAA AGCATAAAAGA GTGACACTTT TTTATTTTA CAAGTGTGTTT AACTGGTTA 360
 ATAGAAGCCA TATAGATCTG CTGGATTCTC ATCTGCTTG CATTCAAGACT ACTGCAATAT 420
 30 TGCACAGAAT GCAGCCTCTG GTAAACTCTG TTGTACACTC ATGAGAGAAT GGGTGAAAAA 480
 GACAATTAC GTCTTACAAT TATTAGAAAT AGCTTTCACT TTAGGAACTC CCTGAGAATT 540
 GCTGCTTTAG AGTGGTAAGA TAAATAAGCT TCTCTTAAA CGGAATCTCA AGACAGAATC 600
 AGTTACATTA AAAGCAAACA AAAAATTGTC CCATGGTTAG TCATCTGTG AAATCTGCCA 660
 CACCTTGGGA CTGGGCTACA ATTGGATAAT ATAGCATTCC CCGAGATAAT TTTCTCTCAC 720
 AATTAAGGAA AGGGCTGAAT AAATATCTCT GTTTGAAGTT GAATAACAAA AATTAGGACC 780
 35 CCCTAAATTT TAGGGCTCCT GAAATTGTC TTTTGCCTA TATTCAAGCTA CTTTACGTT 840
 TATTAATCTCT TCTTCAGGC CAGGTGCCT AGCTCATGCC TAGAATCTCA GGCAGGCCCT 900
 AGCCCAGGAA TTGAGACCA GCCAGGGCAA CACAGTCTCT AAAAAAAAT AAAAAATTAC 960
 CTGGGTGTCT TGGTGCATGC CTGTAGAACT ACTCAGGATG CTGAGGACTG CTTGAGGCCA 1020
 GGATAGCCAA ATCTGTGGTG AGTTCAAGCC CAAACAGAG CGAGACTTT TCAAAAAAAC 1080
 AAACAAAAAA ACAACAAAC TTCCCTCAAA ATAACCTTTT ATCTGCAATG TTTTCCTATT 1140
 40 GCCTGTGAGA TTAATTAC TCTTTTACCT GATTTCAAA GCCCTCCATA ATCTAATCCG 1200
 ACTTTACCTT GTGTTCACTG CAAAATAGCA GGACTGTTCC ACTACAATCC AAAAAATCACA 1260
 GGTTGGGTGC AGTGGCTCAC TCCGTAAATC CCAACACTTT GGAAGGCCAA GGCAGGTGGA 1320
 TTGCTTCAGC TCAGGAGTTTC AAGACCAAGCC TGGCAACAT GGCAAAACCT CTGCTCTCC 1380
 AAAACATACA AAAATTAGCC AGATGTGGTA GTATGTGCCCT GTAGTCCCAA CTACTCAAA 1440
 45 GGCTAAGGCA AGAGGATCAC TTGAGGCCAG GAGGTCAAGG CTACAGTGAG CCATGTTTAC 1500
 TGTGTCAGT CACTCCAGCC TGGGTGATAG AGCAAGACCA TGTCTCAAAA AAAAAAAA 1560
 GAAAAAGAAA GAAAAAAACA TCGCTCTATT CAGTCACCC CCACCCACAAC ATTGTTTGA 1620
 TTATCACATA AATGCTGGTC CATTGCCCTC TCTATCTATT CAAATCTTTA AGCATTCTT 1680
 GAGATTCAAC TCAATTCTCC TTTTCAAAC AGGCCATTAA AACTACATCA GTTCCATTTT 1740

50

55

5 GATTTTCTTG CTTTGAGTCT ACAGACTCAA AAACAAAAAC TTAAAAAACTT ATTTTTAAG 1800
 TTTTCTGCTA CTCTCACTTC TTCAACACTC ACATACACGC ATTACATAATA AGATGGCAGA 1860
 ATGTTCAAGG ATAAAATGAT TTATAGAACT GAAAAGTAG GTTTGATCT TGTTGCTGTC 1920
 AAGATGACTA CCTACCTGAT CTCAGGTAAT TAATTATGTA GCATGCTCCC TCATTTCATC 1980
 CCATACCTAT TCAACACGGAT TGGAATTCCA CAGCAAGGAT AAACATAATC ATAGTTGCTT 2040
 TTCAAGTTCA AGGCATTTA ACTTTAATC TAGTAGTATG TTTGTTGTTG TTGTTGTTGT 2100
 TTGAGATGGA GCCCTGCTGT GTCACCCAGG CTGGAGTGCA GTGGCACGAA CTCGGCTCAC 2160
 TGCAACCTCT GCCTCATGGG TTCAATCAGT TATTCTGCCT CAGTGTCCC AGTAGCTGGG 2220
 10 ACTACAAGGC ACATGCCACC ATGCCGGCT AATTTTGTA TTTTTAGTAG AAACAGGGCT 2280
 TCACCATGTT GCCCAGGCTG GTCTCGAACT CCTGACCTCA AGTGATCCAG CCGCCTCGGC 2340
 CTCCCAAAGT GCTGGGATTA CAGGCATAAG CCACCGTGC CAGCCTAATA GTATGTTTT 2400
 AAACCTTAG TGCGTTAACATGCTGGTT TATAATAAAT ATGCCATAAA TATTTACTGT 2460
 CTTAGAATTATGAGAAGTG GTTACTAGGC CGTTTGCCAC ATATCAATGG TTCTCTCCTT 2520
 ACAGCTTAA TTAGAGTCTA GAATTGCAGG TTGGTAGAGC TGGAACAGAC CTTAAAGATT 2580
 15 GACTAGCCAA CTTCCTGTC CAAATGAGGG AACTGACACC CTTAAAATTAA AGTGACTTGC 2640
 CCCAGACAAA ACTGGAACCT ATGTGTCCTA ATTTCCATCA TGAAATTCTA CCATTCACTA 2700
 GCCTCTGGCT AGTTGTCAAA GTATTGCATA ACTAAATTTC TATGTCTGTT TTAAAGAAC 2760
 AATTGTCACT GCTTACTCCT GGGAGGGTCT TTCTGAGGTG GTTTATAACT CTTAAAAAAA 2820
 AAAAAGTCAG TAGTCTGAGA ATTTAGACG AAATAGTCAA AGCATTTTA TCCAATGGAT 2880
 20 CTATAATTTCATAGATTAG AGTTAACATCA AAGAAACACG GATGAGAAAG GAAGAGGAAA 2940
 ATTGAGGAGA GGAGGAATGG GGATGAGAAC ACACTACTTG TAATCAGTC TAGATGTACT 3000
 GAGAACTAAC AAGAAGAAC CTAAGAAAAT AAGAATGAAG AATTCAAAAT CAACACATGA 3060
 AATAAAAAGA AACTACTAGG GAAAATGGA GAAGACATTA GAAAATTAT TCTATTTTA 3120
 AAATTCTGTT TTCAGGCTTC CCTCCTGTT TCATGGTT TCAGGTTGGAG 3180
 GGAAAGTTA AGATGGAAA AATATATATA TTCTACACAT CCCTTCTAC GCTGTTGTC 3240
 25 TGGCAACAAAG GTTTATCATA GCAAACCTTT ATTACATACAA CATTATTGAT GTTCTTACTG 3300
 TGTGTAAGC TCTTCCAGG TGTTGAAAAT TCAGGGAAA AAAGACAAC CATTGTCCTA 3360
 AAACTCAGAT GAAAGCTGAA CAGACCTATT TTTAATCAAA GAAATCTCAA TTAGGGTAG 3420
 TAAGAGCTAT TTAAGAAGCA TGAACAGGTG TGAAGGAGGT AGGACTCTGA GGAGAGAATA 3480
 GTTAGCTAGG AATGAAAGAG CAGAGAAGTT TTCTCTAGAGG AACTATTAA GCTGGGAGTT 3540
 30 ACGGGATGAA AGATGAGGCAG GGGTTGCAG GCAAAAAAAA AAAAAAGGCA GGGGAAGGGG 3600
 AAGTTCTGGC CTGGCAGAGA GAATAACTGT GGCACAAATG GAGGAGAGTC TGGAAGCAAG 3660
 AAAACCAAGT AGAAGAGTAT TAAAATAGAA GATGCCAGGG GAAATGAGGG CTTGATTAA 3720
 AACAGTGCTG TTGGAGATGG AGAGGAGATA CAAATTCTG GAGACATTTC TGAGTTAGAA 3780
 CCTACAGTAT TTACAGACA AGGGAAAGAT TAGACAAAGG AGTTAAGAAT GACTCCAGG 3840
 35 TTTCAGTTTG GGGCAGGTAA CTAGGACATG TTTTGAAGG TAATGTTATG GATCTCTTAC 3900
 CATTGGAACATGTTAGTCAAGGAAATTA AAATTTGAC ATGTATATAA CTCTCCCCCCC 3960
 ACCACCAAGTA ACTACTTCCC TAACTCTCTA CTTTGATGCC AGACTTCCCTA AAAGAATAGT 4020
 TTGTAGTCAC TGTCTTTACT TTTCCCTCC CATTCTGTC TAGATATTG TCCACCTTAC 4080
 ATCTGCTGCC TCCACTTAC CCAAACGTGTT CTACGGTTGC CCAAACCTTC CTAATTGCCA 4140
 AATTCAATGA ACAAGTTAA GCTTATATGT AAATTAGGAG CTCTACAGTT TGATTTGAG 4200
 40 CAGCCCCCTCC TGAAACCCCTT TCTCTTCGA CTTCTGTGAC ACATCTCAGA TTTACAAAAC 4260
 TGAACATAATT ATTTTACACT TGAGCTGTAT TTCTGTTCTT CTTCTTGTAT GAATGAGGTA 4320
 ACCACTCAAC AAATTGCCCA AGCCAAAAAC TACGAAGTCA TCCTCAGTTC CTCCTCTTC 4380
 TGTGTTGACCC ACAACAGATC AGCTGAGAAA TCCCGCTGTT TAGTATCTCT GAAATTCTATT 4440
 ACCTTAATTAT ATAGCCTCAT CAACTCTAA TTGTTAAAAT TACTTCAGTA GTTGTGTTCT 4500
 GACCTCTGTC CAATCTGTT CAATCAGGTC CATTCTTTG TTCTGGTGG TGGTGGTGGT 4560
 45 GTTGACAGAG TTTCGCTTTT GCTGCCAGG CTGAAGTGCA GTGGAGCACT TCACTGCAAC 4620
 CACAGCTCC TGGGTTAACAGTTACCC TCCCGAGTAG CTGGGACTAC AGGTATGTGC 4680
 CACCAACCCC AGCTAATTTC GTGTTTCAG TAGAGACAGG GTTTCACCAT GTTGGTCAGG 4740
 CTGGTCTCAA ACTCCCTGACC TCAAGCAATC CACCCACCTC AGCCTCCCAA AGTGCTGGGA 4800
 TTACAGGCAT GAGCCACTGC ACACGGACCA GATCCATTGT TTATGTTGCT TCTAGAGTGA 4860
 50 GTTTTAAAAA CACAAATTG ACCATATCTT TCTCCAATT AAGTCAGTAT TTTTTTTTC 4920
 AGGAAAAAAAC AGTTCAAAC TTTAGTCTG CTTACACAAAG GCCTTGTAG TCTGACTCTT 4980
 CTTTCCAAGC TTTCATCAA GTATACTGCA AGTTACATT TATGTGAATT GAATTAGGCA 5040
 ACGGTATAAA AATTATAGTT TATATGGCA AAATGGAAT AATGTTAATC CTTCAAATA 5100
 GTTTATCTAG AATGACATAA TTTCAAAGCT GTCAGGTCAA ATGAGTTATA AACTGTTAAC 5160

5	ACTATTGCCA CATGCAAGTG TCTCTTATAC TTGGTAGAAT TATCTGCTTC CATGTCATTA 5220
	TTATGTAAT TAGACTTTAA ATAACTCAGA AGTTCTTCAG ACATACAGGT TATTATTGTG 5280
	CTTTTAAAC ATAATTTAA ATAATTTAT ATATGATAAT GTTATCCAAG TGCTAAGGCA 5340
	TGTATTGTTA CTGCTGTGCA AAAAAAAA AAAAAAAAAC TCCAAATAAA TATGTTGAAA 5400
	CCAAGTAT ATGCAAGAAA ACAATATTAAC AAAGGCCAA GTACCACCAT AATAGGCTCT 5460
	GTGGAGACGG CAGGCTACAA AACACTAGTA ATAATGCTGA GAAAGTTGAA AAAAGAAAGA 5520
	AAGCAACAAT ATGCTTTGGT TGTTGTAGGT TTATGTACTC CAAGAATATC TCCTCTCAA 5580
	CTTTTACGTT TTTTCCAAG AAAAGTTAAC TTTGGCTGGG CGCAGTGGCT CTTGCCTGTA 5640
10	GTCCCAGCCT TTGGGAGGCC AAGGCCGGCA GATCACCTGA GGTCAGGAGT TTGAGACCAAG 5700
	CCTGACCAAA AATGGAGAAA CCCGCCCCC TCACTACTAA AAGGATACAA ATTAGGCCG 5760
	GGCACAGTGG CTTACCCCTG TGATCCCAGC ACTTTGGGAG GCCGAAGCAG GAAGATCAC 5820
	TGAGGTCAAG AGTCGAGAC CAGCCATGGA GAAACCCGTC TCTACTAAAA ATACAAAATT 5880
	AGCCGGGCGT GGTGGTGCAT GACTGTAATC CCAGCTACTC AGGAGGCTAA GGCAGAGAAT 5940
15	CACTTGAACC CAGGCAGTGG AGGTTGCAGT GAGCCGAGAT CGTCCATTG CACTCCAGCC 6000
	TGGGCAACAA GAGCGAAACT CTGTATCCAA AAAACAAAAG AAAAGAAAAG GTAACCTGTA 6060
	ACTATGTGAG ATCTTTAGAA ATGCATTCTT TCTGTAAAAT GTGACTACAT TTGCCTTATT 6120
	TATGCTAAAAT ATGTTGAGGC CTCAAACAAC CCATATTTC TCGGTCTCCC CGCTGCCTAG 6180
	CCTTGTTCAT CATTGCTCT TCTTGGTGGAGCTCTTCTTCTGTA AAATGCCTGC 6240
	TTCTCTTCA AGGTAGCACA GTCATCACTT TCTGTGGTAA CCTTCTCCAG CACCATCAAA 6300
20	CAGAAAAGAT GAATCTCTTG TAAATTCAAGC TCTTACGTCA TTCATTACAT TATTTTGTA 6360
	CTCTTTATAG ATTCTTCTCT CCCACTAGAC TCTGAGTCAC TGGAGAGTAG GAGCCAAC 6420
	TCATTCTATGT GTGGTTTGGT CAGCTACTGG CCACATTCTT GATGCTAGT TAATGCTCAA 6480
	ACCTTAACGT GTGAATCAGC TCAAATATTG TCCTTCTCTA AATCCATTCA CTCATTGACT 6540
	AACTATGTAC TCAAATAGT AAACACCAAGT AATTAACTCC AATTCTGCC CATACTGCTT 6600
	GGTACATTTG AGGTGAATTA GTTTGATAAA TATGTGTGTA TTACATAATA TTAAAGTATG 6660
25	TACAGAAAGAT CATGCTAACAT ATAATTCAAC ACTGATAACT AATCAACAT AAATGCTCTC 6720
	AGGTTAACAA ATGTCTGCCCT TCTCAGTTAA TGCACTCATT AACAAACACC TTCTGATGCT 6780
	GATAATAGGG CCTTGTTCAG CAATGAAGCC ATAAAGGTGA ATAAAGAACATA TGCCTCGTG 6840
	GAGCTCACAG CCTAGTCATT ATTGTTCTGA TTTTTAATAT TAATGTTGGT TTGGGTTTG 6900
	GTGAAAATG TTTAGACTTA TCTTAGTGAT CTTTCATCC TTTGCTATAT TATTTTCTC 6960
30	TAAGAGTCTT CCTTATCCCC TCCTTAAAAA AACTAGGTGA TAATTCTAAA TTGTAATTTT 7020
	AAATATTATA AATAGCTTAT AAAATTAAAT ATTATATAATA TTTAAATGTT TGATAAAATAT 7080
	TTAAATTTTA TAATATTAA ATGTTTATTAA AATTCAATT GTACATCAGT TTTTATTTTA 7140
	TTTAAATGTG TTGGCCAGGC ATGGTGGCTG ACACCTATAA TCCCAGAACT TTGACAGGCC 7200
	AAGTCAGGCA ACCATTGAG GCTCAGGAGT TTGAGACCAC CCTGGCAAC GTGGTAAAC 7260
	CCTGCTCTCA CCAAACATAT GAAAACCTTAT CTGGGTGTGG TGGCACGCAT CTGTGGTCCC 7320
35	AGATGGGAGT CCCAGGCTAA GATGGGAGAA TCGCTTGAAC CCAGGTGAGA GGGGTGGGGT 7380
	GGATGTTGCA GTGAGCTGAG ATCGTGCAC TCCACTCCAA CCTGGGTGAC AGAGTGAGAC 7440
	TCCATCTCAA AAAAAAAAATGTTATCTAA ATAAGATAAA TTTAATAACT GTTCCGACTT 7500
	AGATGAGCAT AAGGAACATAACCTAGATAA AACTATCAA TAAGGCTGG GTACAGTGAC 7560
	TCATGCTGT AATCTCAAGC ACTTTGGGAG GCCAAATTA TACAAAGTTA GTTGATAAAC 7620
	ACCAACTAAC AACTATTTG GGGTTAGCTT AATTCAAGATT AATTTTTTT AAACCTGAGTT 7680
40	TTAAATTCTC GCTTACTCTA CCATACATGC TAGGCTCTCAT ATTATGCTAG AAAAATTGG 7740
	AGCACAGATT TATGAATACT CTCCTGCATA CCATTTAAAT TTTAAACAAA TTTAATGCA 7800
	GTATATATGT GCCTTTTAC CAACACATTA AATAATAAGA TCTACTGTGA GGACTAAATT 7860
	TCTGTAATT TCAAAGTAGTA ATGAGTTAA ACCATGTCTC AAGATCTCTG CAATAACTGT 7920
	AGCACAACAG AAAATAGGTA TTTCTATTAA TGACAGAGTC ACAAGTACTA CTAATAATAC 7980
	TGTGGTTGT TTCTGCAAC TAATCATGGG AGGAATGCTA AATTTCAGAG GTTGGTGA 8040
45	ATACATGTGT ATTTTTTCC CCACTTCAAGT TCACAGATTT CTCACACTGA GAACTCCAT 8100
	TCCATAACAA AATTCTGGAA GCCTGCACAC CGTATTGGAA GAAGGGCAGA AAGGAAAAGC 8160
	AAATGGAAGG ATTTAAATT TTTCAAATC CTGTATCCCT TGATTTACA GCAAGATTGT 8220
	ATTTATGTAT TACTTGTGTT AAAATATAG TATAATCGAG ACTCCAGATC AAAAATCACC 8280
	GCAGCTCAGG GAGAAAGAGG GCCACCAAAT GCCAGAGCCC TTCAGCCTTC TCCCACCTG 8340
50	CCTGTACCCCT CAGATGGAAG CACTTTTTA TCATTGTTTC ACCTTAGCA TTTTGACAAT 8400
	GAAGTCACAA ACCTTCAGCC TCTCACCCAT AGGAACCCAC TGTTGTAAG AGAAGGATGA 8460
	AGCCACTCCT TCCTAAAGGG CACGATTAGA TGTGTTATG GCATCCTCAG GTCAAACATAT 8520
	ATTTATATTG ACAATATATT TATATTCTC AAGGAATACT AGAATAATGA TTCAGTCAG 8580

5 TACTAGGCCA TTTATCTACC CTTTATAATA TTGTTTAATG AGAAAATGCT TTCTATCTTC 8640
 CAAATATCTG ATGATTTGTA AGAGAACACT TAAACATGGG TATTCTATAAG CTGAAACTTC 8700
 TGGCATTATG TGAATGTCAA GATTGTTCAT CAGTATACTA GGTGATTAAC TGACCACTGA 8760
 ACTTGAAGGT AGTATAAAAGT AGTAGTAAAA GGTACAATCA TTGTCCTCTTA ACAGATGGCT 8820
 CTTTGCTTTC ATTAG 8835

SEQ ID NO: 11:

10 GTAAGGCTAA TGCCATAGAA CAAATACCAAG GTTCAGATAA ATCTATTCAA TTAGAAAAGA 60
 TGTTGTGAGG TGAACATATTA AGTGACTCTT TGTGTCACCA AATTTCACTG TAATATTAAT 120
 GGCTCTTAAA AAAATAGTGG ACCTCTAGAA ATTAACCACA ACATGTCCAA GGTCTCAGCA 180
 CCTTGTCACTA CCACGTGTCC TGGCACTTTA ATCAGCAGTA GCTCACTCTC CAGTTGGCAG 240
 15 TAAGTGCACA TCATGAAAAT CCCAGTTTC ATGGGAAAAT CCCAGTTTC ATTGGATTTC 300
 CATGGAAAAA ATCCCAGTAC AAAACTGGGT GCATTCAAGGA AATACAATTT CCCAAAGCAA 360
 ATTGGAAAT TATGTAAGAG ATTCTCTAAA TTTAGAGTTC CGTGAATTAC ACCATTTAT 420
 GTAAATATGT TTGACAAGTA AAAATTGATT CTTTTTTTTT TTTCTGTG 480
 AGTGCAGTGG CACAATCTCT GCTCACTGCA ACCTCCACCT CCTGGGTTCA AGCAATTCTC 540
 CTGCCTCAGC CTTCTGAGTA GCTGGGACTA CAGGTGCATC CCGCCATGCC TGGCTAATT 600
 20 TTGGGTATTT TTACTAGAGA CAGGGTTTTG GCATGTTGTC CAGGCTGGTC TTGGACTCCT 660
 GATCTCAGAT GATCCTCCCTG GCTCGGGCTC CCAAAGTGT GGGATTACAG GCATGAACCA 720
 CCACACATGG CCTAAAAAATT GATTCTTATG ATTAATCTCC TGTGAAACAAT TTGGCTTCAT 780
 TTGAAAGTTT GCCTTCATTT GAAACCTTC AAAAAAGCC TCAGCAACAA AGTGAGACCC 840
 CATCTCTACA AAAAATGCA AAATATCCTG TGGACACCTC CTACCTCTG TGGAGGCTGA 900
 AGCAGGAGGA TCACTTGAGC CTAGGAATTG GAGCCTGCAG TGAGCTATGA TCCCACCCCT 960
 25 ACACCTCCAGC CTGCATGACA GTAGACCCCTG ACACACACAC AAAAAAAAACCTTCATAA 1020
 AAAATTATTA GTTGACTTTT CTTAGGTGAC TTTCCGTTA AGCAATAAT TTAAAAGTAA 1080
 AATCTCTAAAT TTTAGAAAAT TTATTTTTAG TTACATATG AAATTTTAA ACCCTAGTT 1140
 TAAGTTTTAT GTCTAAATTAA CCTGAGAACAA CACTAAGTCT GATAAGCTTC ATTTATGGG 1200
 CCTTTGGAT GATTATATAA TATTCTGTG AAAGCCAAGA CAGACCCCTA AACCATAAAA 1260
 30 ATAGGAGTTC GAGAAAGAGG AGTAGCAAAA GTAAAAGCTA GAATGAGATT GAATTCTGAG 1320
 TCGAAATACA AAATTTACA TATTCTGTGTTT CTCTCTTTT CCCCCCTCTTA G 1371

SEQ ID NO: 12:

35 GTAAAGTAGA AATGAATTAA TTTTCTTTG CAAACTAAGT ATCTGCTTGA GACACATCTA 60
 TCTCACCATT GTCAGCTGAG GAAAAAAAAT AATCGTTCTC ATGCTACCAA TCTGCCTTC 120
 AAGAAATGTG GACTCAGTAG CACAGCTTTG GAATGAAGAT GATCATAAGA GATACAAAGA 180
 AGAACCTCTA GCAAAAGATG CTTCTCTATG CCTTAAAAAA TTCTCCAGCT CTTAGAAATCT 240
 ACAAATAGA CTTTGCTGT TTCATTGGTC CTAAGATTAG CATGAAGCCA TGGATTCTGT 300
 TGTTAGGGGA GCGTTGCATA GGAAAAAGGG ATTGAAGGCAT TAGAATTGTC CAAATCAGT 360
 40 AACACCTCCT CTCAGAAATG CTTTGGGAAG AAGCCTGGAA GGTTCCGGGT TGGTGGTGGG 420
 GTGGGGCAGA AAATTCTGGA AGTAGAGGAG ATAGGAATGG GTGGGGCAAG AAGACCACT 480
 TCAGAGGCCA AAAAGCTGAAA GAAACCATGG CATTATGAT GAATTCAGGG TAATTCAAGA 540
 TGGAAAGTAGA GTAGGAGTAG GAGACTGGTG AGAGGAGCTA GAGTGATAAA CAGGGTAG 600
 AGCAAGACGT TCTCTCACCC CAAGATGTGA ATTGGACT TTATCTTGA GATAATAGGG 660
 45 TTAATTAAAGC ACAATATGTA TTAGCTAGGG TAAAGATTAG TTTGTTGTA CAAAGACATC 720
 CAAAGATACA GTAGCTGAAT AAGATAGAGA ATTTTCTCT CAAAGAAAAT CTAAGTAGGC 780
 AGCTCAGAAG TAGTATGGCT GGAACCAACC TGATGATATT GGGACCCCCA ACCTTCTTC 840
 GTCTTGTACC CATCATCCCC TAGTTGTTGA TCTCACTCAC ATAGTTGAAA ATCATCATAC 900
 TTCCCTGGTT CATATCCCAG TTATCAAGAA AGGGTCAAGA GAAGTCAGGC TCATTCCCTT 960
 CAAAGACTCT AATTGGAAAGT TAAACACATC AATCCCCCTC ATATCCATT GACTAGAATT 1020
 50 TAATCACATG GCCACACCAA GTGCAAGGAA ATCTGGAAAA TATAATCTTT ATTCCAGGTA 1080
 GCCATATGAC TCTTTAAAT TCAGAAATAA TATATTTTA AAATATCATT CTGGCTTTGG 1140
 TATAAAGAAAT TGATGGTGTG GGGTGAGGAG GCCAAAATTA AGGGTTGAGA GCCTATTATT 1200
 TTAGTTATTA CAAGAAATGA TGGTGTGATG AATTAAGGTA GACATAGGGG AGTCCTGATG 1260

5 AGGAGCTGTG AATGGATTT AGAAACACTT GAGAGAATCA ATAGGACATG ATTTAGGGTT 1320
 GGATTTGGAA AGGAGAAAGAA ACTAGAAAAG ATGATGCCA CATTTCAC TTAGGCAATT 1380
 TGTACCATTC AGTGAATAG GGAACACAGG AGGAAGAGCA GGTTTGGTG TATACAAAGA 1440
 GGAGGATGGA TGACGCATT CGTTTGGAT CTGAGATGTC TGTGGAACGT CCTAGTGGAG 1500
 ATCTCACAA ACTCTTCTAC ATGTGTTCT GAGTTCAGGA CACAGATTG GGCTGGAGAT 1560
 AGAGATATTG TAGGCTTATA CATAGAAATG GCATTTGAAT CTATAGAGAT AAAAAGACAC 1620
 ATCAGAGGAA ATGTGTAAAG TGAGAGAGGA AAAGCCAAGT ACTGTGCTGG GGGGAATACC 1680
 10 TACATTAAA GGATGCAGTA GAAAGAAGCT AATAAACAAAC AGAGAGCAGA CTAACCAAAA 1740
 GGGGAGAAGA AAAACCAAGA GAATTCACC GACTCCCAGG AGAGCATTTC AAGATTGAGG 1800
 GGATAGGTGT TGTGTTGAAT TTTGCAGCCT TGAGAATCAA GGGCCAGAAC ACAGCTTTA 1860
 GATTTAGCAA CAAGGAGTTT GCTGATCTCA CTCAAAACAG CTTGATGGTG AAATGGAGGC 1920
 AGAGGCAGAT TCCAATGAGT GAAACAGTGA ATGGGAAGTG AAGAAATGAT ACAGATAATT 1980
 CTTGCTAAAA GCTTGGCTGT TAAAAGGAGG AGAGAAACAA GACTGCTGC AAAGTGAGAT 2040
 15 TGGGTTGATG GAGCAGTTTT AAATCTAAA ATAAAGAGCT TTGTCCTTT TTGATTATGA 2100
 AAATAATGTG TTAATTGTAA CTAATTGAGG CAATGAAAAA AGATAATAAT ATGAAAGATA 2160
 AAAATATAAA AACCAACCCAG AAATAATGAT AGCTACCATTT TTGATACAAT ATTTCTACAC 2220
 TCCTTCTAT GTATATATAC AGACACAGAA ATGCTTATAT TTTTATTAAA AGGGATTGTA 2280
 CTATACTAA GCTGCTTTT CTAGTTAGTG ATATATATGG ACATCTCTCC ATGGCAACGA 2340
 20 GTAATTGCA GTATATTAAG TTCATGATAT TTCACAATAA GGGCATATCT TTGCCCTTT 2400
 TATTTAATCA ATTCTTAATT GGTGAATGTT TGTTTCCAGT TTGTTGTTGT TATTAACAAAT 2460
 GTTCCCATAA GCATTCCCTGT ACACCAATGTT TCACACATTG TTCTGATTTT TTCTTCAGCA 2520
 TAAAACCCAG GAGGTAGAAT TGCTGGTTG ATAGAAGAGA AAGGATGATT GCCAAATTAA 2580
 AGCTTCAGTA GAGGGTACAT GCCGAGCACA ATGGGATCA GCCCTAGATA CCAGAAATGG 2640
 CACTTCTCA TTTCCTCTTG GGACAAAAGG GAGAGAGGCA ATAACGTGTC TGCCAGAGTT 2700
 25 AAATTGTAC GTGGACTAGC AGGAAATCAT TTGCTGAAAA TGAAAACAGA GATGATGTTG 2760
 TAGACCTCCT GAAAGAGAGCA AAGAAAATTG GAAATTGCGG CTATCAGCTA TGGAAGAGAG 2820
 TGCTGAATG GAAAACAAAAA GAAGTATTGA CAATTGGTAT GCTTGTAAATG GCACCGATTT 2880
 GAACGCTTGT GCCATTGTTG ACCAGCAGCA CTCAGCAGCC AAGTTGGAG TTTTGTAGCA 2940
 GAAAGACAAA TAAGTTAGGG ATTAAATATC CTGGCCAAAT GGTAGACAAA ATGAACCTCG 3000
 30 AGATCCAGCT GCACACGGAA GGAAGGGAG ACGGGAAGAG GTTAGATAGG AAATACAAGA 3060
 GTCAGGAGAC TGGAAGATGT TGTGATATTG AAAAACACAT AGAGTGGAC TAAAAGTGT 3120
 AGAAAACATG AAGGGTAAGA GACCAGTCAG AAAGTAGGCT ATTTGAACCTT AACACTTCAG 3180
 AGGCAGAGTA GTTCTGAATG GTAACAAGAA ATTGAGTGTG CCTTGGAGAG TAGTTAAAA 3240
 AACAAATAGGC AACTTTATTG TAGCTACTTC TGGAACAGAA GATTGTCATT AATAGTTTA 3300
 GAAAACAAAA ATATATAGCA TACTTATTG TCAATTAAACA AAGAAACTAT GTATTTTAA 3360
 ATGAGATTG ATGTTTATTG TAG 3383.

35

40 11. The genomic DNA of claim 9, which comprises additional three introns with respective nucleotide sequences given in SEQ ID NOs: 10 to 12.
 12. The genomic DNA of claim 1, which has a nucleotide sequence selected from the group consisting of SEQ ID NO: 13 and its complementary sequence;

45

SEQ ID NO: 13:

AAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA
 Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala

48

50

55

	-35	-30	-25			
5	ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G		GTAAGG CTAATGCCAT		98	
	Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala					
	-20	-15	-10			
10	AGAACAAATA CCAGGTTCAAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAACT			158		
	ATTAAGTGAC TCTTGTCG ACCAAATTTC ACTGTAATAT TAATGGCTCT TAAAAAAATA			218		
	GTGGACCTCT AGAAAATTAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCACGT			278		
	GTCCTGGCAC TTTAATCAGC AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA			338		
	AAATCCCACT TTTCATGGGA AAATCCCACT TTTCATTGGA TTTCCATGGG AAAAATCCCA			398		
15	GTACAAAATC GGGTCATTC AGGAAATACA ATTTCACAAA GCAAATTGGC AAATTATGTA			458		
	AGAGATTCTC TAAATTAGA GTTCCCGTGA TTACACCATT TTATGTAAT ATGTTTGACA			518		
	AGTAAAAAATT GATTCTTTT TTTTTTTCT GTTGCCTCAGG CTGGAGTGCA GTGGCACAAT			578		
	CTCTGCTCAC TGCAACCTCC ACCTCCTGGG TTCAAGCAAT TCTCCGTGCT CAGCCTCTG			638		
	AGTAGCTGGG ACTACAGGTG CATCCCCCA TGCCTGGCTA ATTTCGGGT ATTTCCTACTA			698		
	GAGACAGGGT TTGGCATGT TGTCAGGCT GGTCTGGAC TCCTGATCTC AGATGATCCT			758		
	CCTGGCTCGG GCTCCCAAAG TGCTGGGATT ACAGGCATGA ACCACACAC ATGGCCTAAA			818		
	AATTGATTCT TATGATTAAT CTCCTGTGAA CAATTGGCT TCATTGAAA GTTGCCTTC			878		
	ATTTGAAACC TTCAATTAAA AGCCTGAGCA ACAAAAGTGAG ACCCCATCTC TACAAAAAAC			938		
	TGCAAAATAT CCTGTGGACA CCTCCTACCT TCTGTGGAGG CTGAAGCAGG AGGATCACTT			998		
20	GAGCCTAGGA ATTGAGCCT GCAGTGAGCT ATGATCCAC CCCTACACTC CAGCCTGCA			1058		
	GACAGTAGAC CCTGACACAC ACACACAAAA AAAAACCTTC ATAAAAAAATT ATTAGTTGAC			1118		
	TTTTCTTAGG TGACTTTCCG TTTAAGCAAT AAATTAAAAA GTAAAATCTC TAATTAGA			1178		
	AAATTATTT TTAGTTACAT ATTGAAATT TTAAACCCCTA GGTTTAAGTT TTATGTCTAA			1238		
	ATTACCTGAG AACACACTAA GTCTGATAAG CTTCATTTA TGGGCTTT GGATGATTAT			1298		
	ATAATATTCT GATGAAAGCC AAGACAGACC CTAAACCAT AAAAATAGGA GTTCGAGAAA			1358		
25	GAGGAGTAGC AAAAGTAAA GCTAGAATGA GATTGAATT TGAGTCGAAA TACAAAATT			1418		
	TACATATTCT GTTCTCTCT TTTCCCCCT CTTAG CT GAA GAT GAT C GTAAA			1470		
	Ala Glu Asp Asp Glu					
	-10					
30	GTAGAAATGA ATTATTTTTT CTTGCAAAAC TAAGTATCTG CTTGAGACAC ATCTATCTCA			1530		
	CCATTGTCAG CTGAGGAAAA AAAAATGG TTCTCATGCT ACCAATCTGC CTTCAAAGAA			1590		
	ATGTGGACTC AGTAGCACAG CTTGGATG AAGATGATCA TAAGAGATAC AAAGAAGAAC			1650		
	CTCTAGCAAA AGATGCTTCT CTATGCCTTA AAAAATTCTC CAGCTCTTAG AATCTACAAA			1710		
	ATAGACTTTG CCTGTTTCAT TGGTCTTAAG ATTAGCATGA AGCCATGGAT TCTGTGTAG			1770		
	GGGGAGCGTT GCATAGGAAA AAGGGATGA AGCATTAGAA TTGTCAAAAA TCAGTAACAC			1830		
	CTCCTCTCAG AAATGCTTTG GGAAGAAGCC TGAAGGGTTC CGGGTGGTG GTGGGGTGG			1890		
35	GCAGAAAATT CTGGAAGTAG AGGAGATAGG AATGGGTGGG GCAAGAAGAC CACATTCA			1950		
	GGCCAAAAGC TGAAAGAAC CATGGCATTG ATGATGAATT CAGGGTAATT CAGAATGGAA			2010		
	GTAGAGTAGG AGTAGGAGAC TGGTGAGAGG AGCTAGAGTG ATAAACAGGG TCTAGAGCAA			2070		
	GACGTTCTCT CACCCCAAGA TGTGAAATTG GGACTTTATC TTGGAGATAA TAGGGTTAAT			2130		
	TAAGCACAAT ATGTATTAGC TAGGGTAAAG ATTAGTTGT TGTAACAAAG ACATCCAAG			2190		
	ATACAGTAGC TGAATAAGAT AGAGAATT TTCTCTAAAG AAAGTCTAAG TAGGCAGCTC			2250		
40	AGAAGTAGTA TGGCTGGAAAG CAACCTGATG ATATTGGAC CCCCCAACCTT CTTCACTT			2310		
	GTACCCATCA TCCCCTAGTT GTTGATCTCA CTCACATAGT TGAAAATCAT CATACTTCT			2370		
	GGGTTCATAT CCCAGTTATC AAGAAAGGGT CAAGAGAAGT CAGGCTCATT CTTTCAAAG			2430		
	ACTCTAATTG GAAGTTAAC ACATCAATCC CCCTCATATT CCATTGACTA GAATTAAATC			2490		
	ACATGCCAC ACCAAGTGCA AGGAAATCTG GAAAATATAA TCTTTATTCC AGGTAGCCAT			2550		
	ATGACTCTT AAAATTCAAG AATAATATAT TTTAAAATA TCATTCTGGC TTTGGTATAA			2610		
45	AGAATTGATG GTGTGGGTG AGGAGGCCAA AATTAAGGGT TGAGAGCCTA TTATTTAGT			2670		
	TATTACAAGA AATGATGGTG TCATGAATTA AGGTAGACAT AGGGGGAGTG TGATGAGGAG			2730		
	CTGTGAATGG ATTTAGAAA CACTTGAGAG AATCAATAGG ACATGATTG GGGTTGGATT			2790		
	TGGAAAGGAG AAGAAAGTAG AAAAGATGAT GCCTACATT TTCACTTAGG CAATTGTAC			2850		
	CATTCACTGA AATAGGGAAC ACAGGAGGAA GAGCAGGTT TGTTGTATAC AAAGAGGAGG			2910		
	ATGGATGACG CATTCTGTT TGGATCTGAG ATGTCTGTGG AACGTCCTAG TGGAGATGTC			2970		
	CACAAACTCT TCTACATGTG GTTCTGAGTT CAGGACACAG ATTTGGCTG GAGATAGAGA			3030		
	TATTGTAGGC TTATACATAG AAATGCCATT TGAATCTATA GAGATAAAAA GACACATCAG			3090		
50	AGGAAATGTG TAAAGTGAGA GAGGAAAGC CAAGTACTGT GCTGGGGGA ATACCTACAT			3150		

	TTAAAGGATG CAGTAGAAAG AAGCTAATAA ACAACAGAGA GCAGACTAAC CAAAAGGGGA	3210
	GAAGAAAAAC CAAGAGAATT CCACCGACTC CCAGGAGAGC ATTCAAGAT TGAGGGGATA	3270
5	GGTGTGTGT TGAATTTCGC AGCCTTGAGA ATCAAGGGCC AGAACACAGC TTTTAGATT	3330
	AGCAACAAGG AGTTGGTGA TCTCACTGAA AGCAGCTTGA TGGTAAATG GAGGCAGAGG	3390
	CAGATGCAA TGAGTAAAC AGTGAATGGG AAGTGAAGAA ATGATACAGA TAATTCTTGC	3450
	TAAAAGCTTG CCTGTTAAA GGAGGAGAGA AACAGACTA GCTGCAAAGT GAGATTGGT	3510
	TGATGGAGCA GTTTAAATC TCAAAATAAA GAGCTTGTG CTTTTTGAT TATGAAAATA	3570
10	ATGTGTTAAT TGTAATTAAT TGAGGCAATG AAAAGAGATA ATAATATGAA AGATAAAAAT	3630
	ATAAAACCA CCCAGAAATA ATGATAGCTA CCATTTGAT ACAATATTC TACACTCCTT	3690
	TCTATGTATA TATACAGACA CAGAAATGCT TATATTTTA TAAAAGGGA TTGTACTATA	3750
	CCTAAGCTGC TTTTCTAGT TAGTGATATA TATGGACATC TCTCATGGC AACGAGTAAT	3810
	TGCAGTTATA TTAAGTTCAT GATATTTCAC AATAAGGGCA TATCTTGC CTTTTATTT	3870
15	AATCAATTCT TAATTGGTGA ATGTTGTT CCAGTTGTT GTGTTATTA ACAATGTTCC	3930
	CATAAGCATT CCTGTACACC AATGTCACA CATTGTCCTG ATTTTTCTT CAGGATAAAA	3990
	CCCAGGAGGT AGAATTGCTG GGTTGATAGA AGAGAAAGGA TGATTGCCA ATTAAAGCTT	4050
	CAGTAGAGGG TACATGCCA GCACAAATGG GATCAGCCCT AGATACCAGA AATGGCACTT	4110
	TCTCATTTC CCTTGGGACA AAAGGGAGAG AGGCAATAAC TGTGCTGCC GAGTTAAATT	4170
20	TGTACGTGGA GTAGCAGGAA ATCATTGCT GAAAATGAAA ACAGAGATGA TGTGTTAGAG	4230
	GTCCTGAAGA GAGCAAAGAA AATTGAAAT TCGGGCTATC AGCTATGAA GAGAGTCCTG	4290
	AACTGGAAA CAAAAGAAGT ATTGACAATT GGTATGTTG TAATGGCACC GATTTGAACG	4350
	CTTGTGCCAT TGTTCACCAAG CAGCACTCAG CAGCCAAGTT TGGAGTTTG TAGCAGAAAAG	4410
25	ACAAATAAGT TAGGGATTAA ATATCCTGGC CAAATGGTAG ACAAAATGAA CTCTGAGATC	4470
	CAGCTCCACA GGGAGGAAG GGAAGACGGG AAGAGGTTAG ATAGGAAATA CAAGAGTCAG	4530
	GAGACTGGA GATGTTGTGA TATTTAAGAA CACATAGACT TGGAGTAAA GTGTAAGAAA	4590
	ACTAGAAGGG TAAGAGACCG GTCAGAAAGT AGGCTATTG AAGTTAACAC TTCAGAGGCA	4650
	GAGTAGTTCT GAATGGTAAC AAGAAATTGA GTGTGCCTT GAGAGTAGGT TAAAAAACAA	4710
30	TAGGCAACTT TATTGTAGCT ACTTCTGAA CAGAAGATG TCATTAATAG TTTTAGAAAA	4770
	CTAAAATATA TAGCATACTT ATTTGTCAAT TAACAAAGAA ACTATGTATT TTTAAATGAG	4830
	ATTTAATGTT TATTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT	4880
	Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu	
	-5 1 5	
	GAA TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC	4928
	Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe	
	10 15 20	
	ATT GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC	4976
	Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp	
	25 30 35	
35	TGT AGA G GTATTTTT TTAATTCGCA AACATAGAAA TGACTAGCTA CTTCTTCCCA	5032
	Cys Arg Asp	
	40	
	TTCTGTTTA CTGCTTACAT TGTTCCGTGC TAGTCCCAAT CCTCAGATGA AAAGTCACAG	5092
	GAGTGACAAT AATTCACTT ACAGGAAACT TTATAAGGCA TCCACGTTTT TTAGTTGGG	5152
40	TAAAAAATTG GATACAATAA GACATGCTA GGGGTCTATGC CTCTCTGAGC CTGCCTTGA	5212
	ATCACCAATC CCTTTATTGT GATTGCTTA ACTGTTAAA ACCTCTATAG TTGGATGCTT	5272
	AATCCCTGCT TGTACAGCT GAAAATGCTG ATAGTTTACG AGGTGTTGGT GCATCTATCT	5332
	GTAATCCTAG CTACTTGGGA GGCTCAAGCA GGAGGATTGC TTGAGGCCAG GACTTTGAGG	5392
	CTGTAGTACA CTGTGATCGT ACCTGTGAAT AGCCACTGCA CCTCAGCCTG GTGTATATAC	5452
	AGACCTGTGTC TCTAAAATTAA AAAAAAAA AAAAAAAAC CTTAGGAAAG GAAATTGATC	5512
45	AAGTCTACTG TGCCTTCCAA AACATGAATT CCAAATATCA AAGTTAGGCT GAGTTGAAGC	5572
	AGTGAATGTG CATTCTTTAA AAATACTGAA TACTTACCTT AACATATATT TAAATATT	5632
	TATTTAGCAT TTAAAAGTTA AAAACAATCT TTTAGAATTG ATATCTTAA AATACTCAA	5692
	AAAGTTGCAG CGTGTGTGTT GTAATACACA TTAAACTGTG GGGTTGTTTG TTTGTTGAG	5752
	ATGCAGTTTC ACTCTGTAC CCAGGCTGAA GTGCAGTGC GTGCAGTGGT GTGATCTCGG	5812
50	CTCACTACAA CCTCCACCTC CCACGTTCAA GCGATTCTA TGCCCTCAGTC TCCCGAGTAG	5872
	GTGGGATTAC AGGCATGCAC CACTTACACC CGGCTAATT TTGTATTTT AGTAGAGCTG	5932
	GGGTTTCACC ATGTTGGCCA GGCTGGTCTC AAACCCCTAA CCTCAAGTGA TCTGCCTGCC	5992
	TCAGCCTCCC AAACAAACAA ACAACCCAC AGTTAATAT GTGTTACAAC ACACATGCTG	6052

5	CAACTTTAT GAGTATTTA ATGATATAGA TTATAAAAGG TTGTTTTAA CTTTTAAATG CTGGGATTAC AGGCATGAGC CACTGTGCCA GCCCTGAAC GTGTTTTAA AAATGTCTGA CCAGCTGTAC ATAGTCTCCT GCAGACTGGC CAAGTCTCAA AGTGGAAACA GCTGTATTAA GGACTATCCT TTGGTTAAAT TTCCGCAAAT GTTCCTGTGC AAGAATTCTT CTAACTAGAG TTCTCATTAA TTATATTTAT TTCAG AT AAT GCA CCC CGG ACC ATA TTT ATT Asp Asn Ala Pro Arg Thr Ile Phe Ile 40 45	6112 6172 6232 6292 6343
10	ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA CGT ATG GCT GTA ACT ATC Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile 50 55 60	6391
	TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile 65 70 75 80	6439
15	ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTTT CAATCATGTT AATATAATCA Ile Ser Phe Lys	6496
	ATATAATTAG AAATATAACA TTATTCTAA TGTTAATATA AGTAATGTA TTAGAAAACT CAAATATCCT CAGACCAACC TTTTGCTCTAG AACAGAAATA ACAAGAACCA GAGAACCAATT AAAGTGAATA CTTAACAAAA ATTATCAAAC TCTTACCTA TTGTGATAAT GATGGTTTT CTGACCCCTGT CACAGGGCAA GAGGAGATAC AACACTTGT TTATGACCTG CATCTCTG ACAATCAGTC TTATACAAA TAATAATGTA GAATACATAT GTGAGTTATA CATTAAAGAA TAACATGTGA CTTCAGGAA TGAGTCTGC TATGAAGAAT GAAGCTAATT ATCCTTCTAT ATTCCTACAC CTTTGTAAT TATGATAATA TTTAATCCC TAGTTGTTTT GTTGCTGATC CTTAGCCTAA CTCTTAGACA CAAGCTTCAG CTTCCAGTTG ATGTATGTTA TTTTAATGTT TAATCTAATT CAATAAAAGT TATGAGATCA GCTGTAAGAAG TAATGCTATA ATTATCTTCA ACCCAGGTAT AAAGTATTC TGGCCTCTAC TTTTCTCTA TTATTCTCCA TTATTATTCT 25	6556 6616 6676 6736 6796 6856 6916 6976 7036 7096 7156 7216 7276 7336 7396 7456 7516 7576 7636 7696 7756 7816 7876 7936 7996 8056 8116 8176 8236 8296 8356 8416 8476 8536 8595 8656 8716 8776 8836 8896 8956 9016
30	CTATTATTT TCTCTATTC CTCCATTATT GTTAGATAAA CCACAATTAA CTATAGCTAC AGACTGACCC ACTAACAGTA GCCAGGGATG CTTACAAATT GGCAATGCTT CAGAGGAGAA TTCCATGTCA TGAAGACTCT TTTTGAGTGG AGATTTGCCA ATAAATATCC GCTTCATG CCACCCAGTC CCCACTGAAA GACAGTTAGG ATATGACCTT AGTGAAGGTA CCAAGGGCA ACTTGGTAGG GAGAAAAAAAG CCACTCTAAA ATATAATCCA AGTAAGAACCA GTGCATATGC AACAGATAAC GCCCCCAGAC AAATCCTCA GCTATCTCCC TCCAACCAGA GTCCCCACCCCC TTCAGGTGAC AATTGGAGT CCCATTCTA GACCTGACAG GCAGCTTAGT TATCAAATA GCATAAGAGG CCTGGGATGG AAGGGTAGG TGAAAGGGT TAAGCATGCT GTTACTGAAC AACATAATTA GAAGGGAAAGG AGATGCCAA GCTCAAGCTA TGTGGATAG AGGAAAACCTC AGCTGAGAG CCAGATTCAAG AAACCTGGAT AAGTCCGAAC CTACAGGTGG ATTCTTGTG AGGGGAACTG GTGAAATGT TAAGAAGATG GAAATAATGCC TTGGCACTTA GTAGGAACCTG GGCAATCCA TATTGGGGG AGCCTGAAGT TTATTCATT TTGATGGCCC TTTAAATAA 35	
40	AAAGAATGTC GCTGGCGCTG GTGGCTCACA CCTGTAATCC CAGCACTTG GGAGGCCAG GGGGCGGGAT CACCTGAAGT CAGGAGTTCA AGACCAGCCT GACCAACATG GAGAAACCCCC ATCTCTACTA AAAATACAAA ATTAGCTGGG CGTGGTGGCA TATGCCCTGTA ATCCCAGCTA CTCGGGAGGC TGAGGCAGGA GAATCTTTG AACCCGGGAG GCAGAGCTTG CGATGAGCCT AGATCGTGC ATTGCACTCC AGCCTGGCA ACAAGAGCAA AACTCGGTCT CAAAAAA AAAAAAAG TGAAATTAAC CAAAGGCATT AGCTTAATAA TTAACTACTG TTTTAAGTA GGCGGGGGC TGGCTGGAAG AGATCTGTGT AAATGAGGGG ATCTGACATT TAAGCTTCAT CAGCATATA GCAATCTGC TTCTGGAAGG AACTCAATAA ATATTAGTTG GAGGGGGGGA GAGAGTGAGG GGTGGACTAG GACCAGTTT AGCCCTTGTG TTTAATCCCT TTTCTGCCA CTAATAAGGA TCTTAGCAGT GGTTATAAAA GTGGCCTAGG TTCTAGATAA TAAGATACAA CAGGCCAGGC ACAGTGGCTC ATGCCATATAA TCCCAGCACT TTGGGAGGGC AAGGCGAGTG 45	
50	TCTCACTTGA GATCAGGAGT TCAAGACCA CCTGGCCAGC ATGGCGATAC TCTGTCTCTA CTAAAAAAAG TACAAAATT AGCCAGGCAT GGTGGCATGC ACCTGTAATC CCAGCTACTC GTGAGGCTGA GGCAGAAGAA TCGCTGAAA CCAGGAGGT TAGGCTGCAG TGAGCTGAGA TCGCACCACT GCACTCCAGC CTGGCGACA GAATGAGACT TTGTCTCAA AAAAGAAAAA GATACAACAG GCTACCCCTTA TGTGCTCACC TTTCACTGTT GATTACTAGC TATAAAGTCC TATAAAGTTC TTTGGTCAAG AACCTGACA ACACAAAGAG GGATTGCTT TGAGAGGTTA CTGTCAGAGT CTGTTTCATA TATATACATA TACATGTATA TATGATCTA TATCCAGGCT TGGCCAGGGT TCCCTCAGAC TTTCCAGTGC ACTTGGGAGA TGTTAGGTCA ATATCAACTT TCCCTGGATT CAGATTCAAC CCCTCTGAT GTAAAAAAA AAAAGAAAAA GAAAGAAATC	

	CCTTCCCCCT TGGAGCACTC AAGTTCAACC AGGTGGGGCT TTCCAAGTTG GGGGTTCTCC	9076
5	AAGTCATTG GGATTGCTT CACATCCATT TGCTATGTAC CTTCCCTATG ATGGCTGGGA	9136
	GTGGTCAACA TCAAAACTAG GAAAGCTACT GCCCAAGGAT GTCCCTTACCT CTATTCTGAA	9196
	ATGTGCAATA AGTGTGATTA AAGAGATTGC CTGTTCTACC TATCCACACT CTCGCTTCA	9256
	ACTGTAACTT TCTTTTTTTC TTTTTTCTT TTTTTCTTT TTTTTGAAAC GGAGTCTCGC	9316
	TCTGTCGCCG AGGCTAGAGT GCAGTGGCAC GATCTCAGCT CACTGCAAGC TCTGCCTCCC	9376
	GGGTTCACGC CATTCTCCTG CCTCACCCCTC CCAAGCAGCT GGGACTACAG CGGCCTGCCA	9436
	CCATGCCAG CTAATTTTT GTATTTTAG TAGAGACGGG GTTCACCAGT GTAGGCCAGG	9496
10	ATGGTCTCGA TCTCCTGAAAC TTGTGATCCG CCCGCCTCAG CCTCCCAAAG TGCTGGGATT	9556
	ACAGGGCTGA GCCATCGCAC CGGGCTCAAC TGTAACCTTC TATACTGGTT CATCTCCCC	9616
	TGTAATGTTA CTAGAGCTT TGAAAGTTTG GCTATGGATT ATTTCTCATT TATAACATTAG	9676
	ATTCAGATT AGTTCCAAT TGATGCCAC AGCTTAGGGT CTCTCCTAA ATTGTATATT	9736
	GTAGACAGCT GCAGAAGTGG GTGCCAATAG GGGAACTAGT TTATACTTTC ATCAACTTAG	9796
	GACCCACACT TGTGATAAA GAACAAAGGT CAAGAGTTT GACTACTGAT TCCACAACTG	9856
15	ATTGAGAAAGT TGGAGATAAC CCCGTGACCT CTGCCATCCA GAGTCTTCA GCCATCTTG	9916
	AAGGATGAAG AAATGCTATT TTAATTCTGG AGGTTTCTCT ATCAGTGCCTT AGGATCATGG	9976
	GAATCTGTGC TGCCATGAGG CCAAAATTAA GTCCAAAACA TCTACTGGTT CCAGGATTAAC	10036
	CATGGAAGAA CCTTAGGTGG TGCCCCACATG TTCTGATCCA TCCTGCAAAA TAGACATGCT	10096
	GCACTAACAG GAAAAGTGCA GGCAGCACTA CCAGTTGGAT AACCTGCAAG ATTATAGTTT	10156
20	CAAGTAATCT AACCATTCT CACAAGGCC CATTCTGTGA CTGAAACATA CAAGAACATG	10216
	CATTGGCCT TCTAAGGCAG GGCCCAGCCA AGCAGACCAT ATTCAAGGACA GAAATTCAAG	10276
	ACTACTATGG AACTGGAGTG CTTGGCAGGG AAGACAGAGT CAAGGACTGC CAACTGAGCC	10336
	AATACAGCG GCTTACACAG GAACCCAGGG CCTAGCCCTA CAACAAATTAT TGGGTCTATT	10396
	CACTGTAAGT TTTAATTCA GGCTCCACTG AAAGAGTAAG CTAAGATTCC TGGCACTTTC	10456
	TGTCTCTCTC ACAGTTGGCT CAGAAATGAG AACTGGTCAG GCCAGGCATG GTGGCTTACA	10516
25	CCTGGAATCC CAGCACTTTG GGAGGCCGAA GTGGGAGGGT CACTGAGGC CAGGAGTTCA	10576
	GGACCAAGCTT AGGCAACAAA GTGAGATACC CCCTGACCCC TTCTCTACAA AAATAAATT	10636
	TAAAAAATTAG CCAAATGTGG TGGTGATATA TTACAGTCCC AGCTACTCAG GAGGCTGAGG	10696
	CAGGGGGATT GCTTGAGCCC AGGAATTCAA GGCTGCACTG AGCTATGATT TCACCACTGC	10756
	ACTTCTGGCT GGGCAACAGA GCGAGACCCCT GTCTCAAAGC AAAAAGAAAA AGAAACTAGA	10816
	ACTAGCCTAA GTTTGTGGGA CGAGGTCATC ATCGTCTTTA GCCGTGAATG GTTATTATAG	10876
30	AGGACAGAAA TTGACATTAG CCCAAAGGTT TGTTGGCTTG TGCTGAACT CTAACCTAAC	10936
	TTGAGCAAAT GTGGACACCA CTCATGGGA GAGGAGAGAA GTAAGCTGTT TGATGTATAG	10996
	GGGAAACTA GAGGCCCTGGA ACTGAATATG CATCCCATGA CAGGGAGAAT AGGAGATTCG	11056
	GAGTTAAGAA GGAGAGGAGG TCAGTACTG TGTTCAGAGA TTTTTTTAT GTAACTCTTG	11116
	AGAAGCAAAA CTACTTTGT TCTGTTGGT ATATACATTC AAAACAAACT TCATATATTC	11176
	AAATTGTTCA TGTCCCTGAAA TAATTAGGTA ATGTTTTTTT CTCTATAG GAA ATG AAT	11233
35		Glu Met Asn
		85
	CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG	
	Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Glu	
	90 95 100	
40	AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA	11329
	Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser	
	105 110 115	
	TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA	11377
	Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys	
	120 125 130 135	
45	CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC	11425
	Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe	
	140 145 150	
	ACT GTT CAA AAC GAA GAC TAGCTATTAA AATTCATGC C	11464
	Thr Val Gln Asn Glu Asp	
	155.	
50		

13. The genomic DNA of claim 1, which has a nucleotide sequence selected from the group consisting of SEQ ID NO: 14 and its complementary sequence;

SEQ ID NO: 14:

5	ACTTGCCCTTA	AAAGCTTTGC	ATAGGTAGAC	AACATTAGAT	TAATTCCCTT	GCTCACATCT	60
	GTTCAAGAAA	AATCATTAA	GTTATAAAAT	ATAACAAACC	TTCTGCATTA	TAAGACTGAT	120
	GTTTAGAAAT	ATAAACATT	TATACATCAC	CATTTAAATC	TTTCTCCAAG	GCTTCATCTT	180
	TATAAAATAG	TCCCGAAATT	TCAGAGAAAG	ATGAATCTGA	TTTTCCAAGA	GAGGACAGCT	240
	GTGGACTATC	TGGCACTGGA	GACTAAATAA	AGAAAGCAGG	TACACTCAAT	AAGATCTTC	300
	GGACATATAC	ATTTTGTTA	TTAAGAAAAA	GCAAATAAAA	CATTTTCAG	AAAAGGCAA	360
	ACATGCTAGA	AAGCATATGA	CTTAGTCATT	TGAGTTTTA	TTATTAAGGA	AATTTACAGG	420
10	CCCAAGAAAAC	ACCTTGCTCA	ATATATTAA	TTTTATTTTG	GTTCACACT	AGACTTTGCT	480
	TTTCATTGT	TTGTTTTGT	GACAAGTTCT	CGCTCTGTCA	CCTAGGCCAA	AGTGTAGTGA	540
	CACAATCTTA	GCTCACTGTA	GCCTCTAGA	TTCAAGTGTAT	CCTCTGTCT	CAGACTCCTG	600
	AGTAGCTAGG	ACTACAGGAA	CATTCCACCA	TGCCCAGCTA	ATTTGTTT	GTTCAGTTT	660
	GTTCAGAG	ACAATGTATT	GCAGCGTTGC	CCAGGCTGAT	CTGAAACTCT	TAGCCTCAA	720
15	CGATACTCCT	GCCTCAGCCT	CCAAAGCAC	TAGGATTACA	GACATGAGCC	AATGCCCCA	780
	GCCTTAATT	AGACTTTAA	TGTGGTTTA	AACTCCTGTT	AAAAAGCGT	CTGGTATCTT	840
	GAACCACTAG	ATGTTTCAT	AGCAATGAAG	CTAAACTGTA	ATTTAGACAG	TAGCCAAATG	900
	CTTGTGAAAT	TTTCTTAAT	AATATAATCT	TCAAGGGAGC	AAATCATGTC	CCAAATGCAA	960
	AAGATCAACT	GGTGGGGGCA	GTAGTAAAG	ACAGGATACT	GTGCTCTTTA	AAAGGTCAGT	1020
	AACTATAGTA	CCTAGTTATC	TTACTTATCA	CAGCAAAATA	ATTACATAAA	ATCCTATGGA	1080
20	TCATAAAGGC	ACAGACTCAC	TTCTGCTCT	AGATCTCAAG	CTACCAAAAA	AAAATCTCCC	1140
	AATAGTTCT	TGGAGGCCTA	TACTTAGTGA	AAAAGCAGCT	GGAATCAACA	TAGTCCCTCC	1200
	TATGTTGTTAG	GACAATCCTA	GCTCTGGCA	TACGAATACA	TTAAATCCCA	CTTATCTATA	1260
	GAGCTTCTT	AAAGGGAAAGA	AATTGAGTA	GTATGTAAAAA	CAGAATAAAA	GATTAAGGCT	1320
	CCATAGGCAT	ACAGCTTACC	TCCAATTCTC	TTGGCCTCTT	GCAATTCTA	TTATCAGGCT	1380
	TTACAAGCTG	ATTGCCATC	ATATTCCGA	GGCACCGCT	ACAAAGCTTA	GAACAATGCC	1440
25	AGATTTAGGT	ACAAACTCCA	TGCTACAAGC	TCTCTGGAT	CCTCCCTGT	TTCCCACTCC	1500
	TACTGCTGAT	GTAAATTAG	ACTGTCATTA	TCTGTCACTT	TCCTAAACTC	AATTCTCCC	1560
	TCCTCTAAAT	CATTCTATCA	ACTGCTATTT	GGGTAAATCTT	TCAAAACTTT	GATTACTGCA	1620
	TTCCCTTAAC	TCAAAAACCTT	TCATTGTTCC	AGAATAAGTT	GAAATTCCAT	GATATGGCCT	1680
	TCAAGGTCCT	GTATTATCTG	GTGCAAGCCT	ACTAGTCCCA	TCATTTCAA	CTACTCCTCT	1740
30	CTATGTACTT	AGCCAAATGA	GTCTCTCTGG	CAATTCTGCC	TTGTTTCAGG	ACTGGCTCAG	1800
	TTAAGATCT	TTTATCTTCG	GCCGGGCGCG	CTGGCTCACG	GCTGTAATCC	CAGCACTTTG	1860
	GGAAGCTGAG	GCAGGAAGAT	CACCTGAGGT	CGGGAGTTCG	AGACCAGCCT	GGCCAGCATG	1920
	GTGAAACCT	GTGTCTACTA	AAAATCCAA	CATTAGCCAG	GCGTGGTGGC	AGGCGCCTGT	1980
	AATCCCAGCT	ACTGGGAAG	CTGAGGTGAG	AGAATCGCTT	GAACCCAGGA	GAGGGAGGTT	2040
	GCAGTGAGCC	GAGATTGTGC	CATTGCACTC	CAGCCTGGGC	AACAGAGCGA	GACTCCACCT	2100
35	CAAAAAAA	AAGGATTCTT	CTATCTTCAC	AAAATCTAA	TGTTTAAACA	GGTCTTACAG	2160
	TTCATCTAAAT	TCATCTCAT	TTTTACAAG	TGAGAAAACA	GGGACAGTGA	CGGTGGATCA	2220
	AGTGACACCA	GTAGACTGA	GCTAAATTAG	AACCGAGATC	TCACTCGAGT	CTGAGGTTAT	2280
	TCCCCTGTC	CAACCTTACT	TTAAAGTAGC	TTCAAATTTT	ACTTTACTT	TTCCATAAAT	2340
	TCGGAAGGGA	TTTCCCTAG	GAGTCCAAAT	GTGAAACCT	GGAAGGCTAT	AGTCTCTGTG	2400
	TCTTGAGAT	GAGGGGAGCC	CTGTCCATAT	TCAAGTTATC	AATTGACTTT	GTTGTTTTG	2460
40	AGAAACGATG	CTGATTGGG	TAACCTAAC	ACATCTGTTT	GATTAGTCCT	ATAAAATATG	2520
	CATATATAGA	AGACAGAAAG	AGCAACAAACA	AATTGAAAG	ATGCTGTTA	AGTAATTCT	2580
	GTATCGTACG	TGTCCTATTCC	TGCCAGTACC	TTTATAGTAT	GTAAGTTAC	GTGCTGTAAT	2640
	AGTATTAAATA	GTATCTAGAA	AATACTACAC	ATGCACAGCA	GTGCTAACTT	TGCTCTGGGA	2700
	GTTGGAAAAT	ACTTCAGAGA	AGCCAACAGG	CAGATTTTC	TCTCTCCCT	TCCCCTCTA	2760
45	ATTTCCTCTT	TCCCCCTCAC	CCCCCTCTCT	TCTCTCCCCA	AGTAACACTG	TGCACCTATG	2820

5	TCAAACGAAA	ACTTATAATC	AAGTAACTGT	TTCTGCAAAA	ATAAGTTCGT	TTTCCTGTCA	2880
	TGGCTCAAGG	CCTCAGCAGA	TCCAGGCCTG	GTGGACGGGC	TGGTCTTCGT	CGTGTGCCAA	2940
	ACACTGACCA	CTGCCCTGGC	TCTGCCATCT	TAGGCTTAGT	GACCTGGCTG	TTACTAAGCA	3000
	CTGTCCTCTC	TGCCCTCATGC	AGCTGTCTCC	TTCTAGTCTT	CTCCCTCTTC	TCAACGCGAT	3060
	CCTAGCCCCCT	CAGCCCCATT	CACCTCCATT	TTCCCTCACT	TCCCCCGGCC	CCTCCCCACT	3120
10	TCCTCCCTAC	TGTTGTTTCC	GCCCCACTAG	AGCCCCCTCA	AGAAAGTTTC	CATCCTCGCA	3180
	CCCTTCCCTG	TGTCACAGCC	CGTCACATTC	TCACAGGCGC	CCATCCCTCC	AGCCCCACCC	3240
	CAAGGCCAAT	GTACTTCGCG	GTATGGGGAC	CTTCCTCGTC	AGCGAACGCG	AGGGACTGAA	3300
	GACCTGGGC	GCAGGGGTGCT	CGGACTTCGG	GGGTGGAGGT	GGGAAGCGCG	CCGCACTCCC	3360
	AGCAGCCCC	GCACGAGTC	CGTGACAGCT	CTCCCACAC	CACCCCCCCC	AACTTCCCCA	3420
	CCGTAGGCC	CCAGAGGCCAG	GCCCCACGGA	AAGGCAGCTT	TTTCCCGGTT	TTCTCCCGCT	3480
	CTTTCCCTC	CACTTGGAA	ACTCGTGA	AAAAAAATCTC	TCCTGCCCC	CCTGTGTGTG	3540
15	TTTGAACCA	GAAAAAAATCT	GAAACTGGTC	AAGAAAGAAC	AAGGAAGACT	TGCCAAAGCA	3600
	AGGCCGGTGT	GTGTCCCAGC	AGCTTAGAAT	CTCAGCAAAG	GAACACAAAAA	TAGCACATCC	3660
	ACGGCCTCTT	TTCGAGTAA	ATTTACTTGG	TTTGTGTTGCA	GGAAAGGTTT	AAAAGTGGT	3720
	TTGCAGATGC	TCTGTTTGCA	GGAAGGCTT	AATCACGTGT	TCCTCTGGCC	CACAAGCAAG	3780
	GCTTTAGAT	CCAGAGCCTC	AGTTACTGCC	CCCTCTTCC	CTTGGTGCA	ACCAAACGTT	3840
	CAGAACATCAG	CCTCTTCTAGA	AAATTCTTAC	CCCGGGGTGTG	TCAATAAGTT	AAAGTCTAATT	3900
20	GGCAACACGCT	ATCAAAAAGT	GTGCAAAAC	ACACATGCC	CACATAATTC	TAGCTT'GCC	3960
	TCATCGGGTG	TTTTAATGCG	GAGGCTTGA	CCTGCAATT	CAAAGATATA	CATTCCAAGC	4020
	TTACGCCAG	TTAGTGGATG	TGGAAGAAAA	AAAAAAGCAA	ATTACCTCAT	ACACAAAGG	4080
	TCAATAACAC	ACATCCATAA	GCTCCAGGT	AAAAATCTTA	CATCTTAGAG	AACTATATT	4140
	AAACATTACA	TACATTACTA	AGGTTTTTT	TTTCCCTTTG	CTTGATTAAA	TGTTAGTTAT	4200
	CATTAAGTCT	TGGAATTATT	CTGTTGCTG	ATATTATT	GCTGTTGTTG	AAGAAGGCCG	4260
25	TTGTTTAA	TAAGTCCCTA	GAAAATAACG	GCTCAATGTG	TTAAATCTGA	GTTGCTAATA	4320
	TTGTGAATA	TAGGCCACAT	AATACTAGCC	TAGATAACTA	TGGCGAAGTA	AGGAGTCTCA	4380
	AACACTGTCC	CAGAACAAATA	GCAATCTGTG	TTGAATT	ACCCCTCTGTG	GTAAAATGAA	4440
	GGGAAAAGGA	ATGAAGTTT	AGTTGCCTT	ATTTTTATC	TTTATTGTTT	CAGACTCTTC	4500
	AGCAGTATAA	AGTTTCATC	AAAGTCAAATA	TATTCACTT	AAAGTGA	TGCTTTATTC	4560
30	TGATACCATC	TCCTCCCTAA	TTTGGGGGC	CAGCTGACAT	AAGTTTATG	AAATAAAAAG	4620
	ATTAAAAATT	CTTACATTTT	TAGTGTCTT	CCTTGGTAAA	ATGTAGAGTT	GTCCACTGTG	4680
	TTTATCTCCT	CCTCCTTATT	ATCATGGTTG	CTGTTATTAT	TTTAATGGT	TCATTAAACC	4740
	CAAGGGCTC	GGAAATACTC	ATGGAATTCA	TCTCACAGCC	TTCACACTGT	ATGATATT	4800
	AACAGGTGGT	TGTCCATCTG	ATTCTAAAA	TATTCCAAG	AAAAATGATT	CCACCTAATG	4860
	CATAAAATGCT	TTCATCAGAT	TAAGAGAAC	CCATGGACAT	TTTATT	TTTATT	4920
35	AAATATTAAC	TTCCATTGCA	TAAGCTAA	GGGTAGGAAT	AAGTGA	GATATTGTTAT	4980
	CTAGAGCTT	AAAATATTCA	AAGGGCTGTC	ATCATTATCT	CATTAAATCT	TTGAAAACAA	5040
	CTCTATGAA	TACAAAGGAC	ACTGAGACAT	TTGTTGCTCT	ATATCAAAGA	AAAAAGTGT	5100
	TGTCCCCAAA	CTTCAAAATG	TGTAATTAC	ACATTCTGCA	TCTTACAGC	TGGAGAAAAT	5160
	TCACTGGCAA	TGGAATTATT	AAAATTAGAG	CTTGCTTAGT	GTGCTGCTTC	TGATCACTAC	5220
	TTGATCCAC	TTCGTGCTT	CATGTTAATT	GGCCCAATTG	GACTCTACAG	TGGAAGGTG	5280
40	AAAACCTACT	ATTCACACT	GAGTCACGTA	TGTATTCTTA	TCATATACT	CTTAAAGGTA	5340
	CTATTTTTT	TCTTCTGATA	GTCACCCAC	CAAGCACTTC	CAGCCACCC	GCCACAGACT	5400
	TCCCTTGTA	TCACTGTTG	AGGACATGAT	TTTTTATGA	CTTCCCGAAA	TGAAAACCT	5460
	ATCTTGT	TAACCAAAAC	AAACCAACAA	AAAGTACTGT	TTATGTAAGC	ATTTTGTTCC	5520
	CTGACTCTAG	GAACCCCTCT	GTTTTATAT	CAACTCTGTA	CTGGCAAAAC	ACAAAAACAA	5580
45	AATGCCACCT	TGCTAATTCC	CTTCCTAGCA	AAGTAATACA	GTTTAGCACA	TGTTCAAGAA	5640
	AAAAATGGCT	AAGAAATTTT	GTTTCCACTA	ATTATTTCA	AGACTGTGAT	ATTTACACTC	5700
	TGCTCTCAA	ACGTTACATT	TTATAAGACT	ATTTTTAAC	ATGTTGAACA	TAAGCCCTAA	5760
	ATATATGTTAT	CCTTAAATTG	TATTTCAAAT	ATTTTAGGT	AGTCTT'GCT	ATCATTCCAG	5820
	GAATAGAAG	TTTAAACACT	GGAAACTGCA	AGTAAATATT	TGCCCTCTTA	CCTGAATT	5880
	GGTAGGCC	TCCCCAAGCT	TACTTTCTGT	TGCAGAAAGT	GTAAAATTA	TTACATAAAA	5940
50	TTCTAATGAT	GGTATCCGTG	TGGCTGCA	CTGATACACC	AGATAAAGAA	GTTTTATGAA	6000
	AATGGACTCC	TGTTCCACTG	AAAAGTAAAT	CTTAATGGCC	TGTATCAACT	ATCCTTTGAC	6060
	ACCATATTGA	GCCTGGGAGG	AAGGGAAAGT	CCTGAATGAG	GTTATAAAGT	AAAAGAAAAT	6120
	ATTTGCAAAA	TGTTCCCTTT	TTTAAATGT	TACATT	AAATATT	ACTGTTGTA	6180
	CATTGTAGGA	ATTACCCCAA	TAGGACTGAT	TATTCCCGCAT	TGTAAAATAA	GAAAAGTTT	6240

	TGTGCTGAAG TGTGACCAAG AAGTCTGAAA ATGAAGAGAG ACAGATGACA AAAGAAGATG	6300
	CTTCTAATGG ACTAAGGAGG TGCTTTCTTA AAGTCAGAAA GAGATACTCA GAAAGAGGTA	6360
	CAGGTTTTGG AAGGCACAGA GCCCCAACTT TTACGGAAGA AAAGATTTCA TGAAAATAGT	6420
5	GATATTACAT TAAAAGAAGT ACTCGTATCC TCTGCCACTT TATTTGACT TCCATTGCC	6480
	TAGGAAAGAG CCTGTTGAA GGCGGGCCCA AGGAGTGCAG ACAGCAGTCT CCTCCCTCCA	6540
	CCTTCTTCCT CATTCTCTCC CCAGCTTGCT GAGCCCTTG CTCCCTGGC GACTGCCCTGG	6600
	ACAGTCAGCA AGGAATTGTC TCCCAGTGC TTTGCCCCCT CGGCTGCCA ACTCTGGCTG	6660
	CTAAAGCGGC TGCCACCTGC TGCACTAC ACAGCTTCGG GAAGAGGAAA GGAACCTCAG	6720
10	ACCTTCCAGA TCGCTCCCTC TCGAACAAA CTATTGTCG CAGGTAAGAA ATATCATTCC	6780
	TCTTTATTTG GAAAGTCAGC CATGCAATT AGAGGTAAT AACCTAGAAA GCAATTGAGA	6840
	GGAATATAAA CCATCTAGCA TCACATACGAT GAGCAGTCAG TATCAACATA AGAAATATAA	6900
	GCAAAGTCAG AGTACAATT TTTCTTTA TCAGATATGG GAGAGTATCA CTTTAGAGGA	6960
	GAGGTTCTCA AACTTTTGCG TCTCATGTT CCTTACACT AACGACATCA CATGTTAGCA	7020
	TAAGTAACAT TTTTAATTAA AAATAACTAT GTACTTTTT AACACACAAA AAAAGCATAA	7080
15	AGAGTGACAC TTTTTTATT TTACAAGTGT TTTAACTGGT TTAATAGAA GCAATAGAT	7140
	CTGCTGGATT CTCATCTGCT TTGCATTCA ACTACTGCAA TATTGCACAG AATGCAGCCT	7200
	CTGGTAAACT CTGTTGTACA CTCATGAGAG AATGGGTGAA AAAGACAAAT TACGTCTTAG	7260
	AATTATTAGA AATAGCTTC ACTTTAGGAA CTCCCTGAGA ATTGCTGCTT TAGAGTGGTA	7320
	AGATAAATAA CCTTCTCTTT AAACGGAATC TCAAGACAGA ATCAAGTACA TTAAAAGCAA	7380
	ACAAAAAAATT TGCCCATGGT TAGTCATCTT GTGAAATCTG CCACACCTT GGACTGGGCT	7440
20	ACAATTGGAT AATATAGCAT TCCCCGAGAT AATTTCTCT CACAATTAAG GAAAGGGCTG	7500
	AATAAATATC TCTGTTGAA GTTGAATAAC AAAAATTAGG ACCCCCTAAA TTTTAGGGCT	7560
	CCTGAAATTGCTCTTTGC CTATATTCACT TACTTTACG TTCTATTAAA TCTTCTTCA	7620
	GGCCAGGTGC ACTAGCTCAT GCCTAGAACAT TCAGGCAGGC CTGAGCCCAG GAATTGAGA	7680
	CCAGCCAGGG CAACACAGTC TCTACAAAAA ATAACAAAT TACCTGGGTG TGTGGGTGCA	7740
25	TGCTGTAGA ACTACTCAGG ATGCTGAGGA CTGCTTGAGC CCAGGATAGC CAAATCTGTG	7800
	GTGAGTTCA GCACTAAACA GAGCGAGACT TTCTCAAAA AACAAACAAA AAAACAAACA	7860
	AACTTCTTTC AAAATAACTT TTTATCTGCA ATGTTTCTC ATTGCTGTG AGATTAATT	7920
	TACTCTTTA CCTGATTTC AAAGCCTCC ATAATCTAAT CCGACTTAC CTTGTGTCA	7980
	CTGCAAAATA GCAGGACTGT TCCACTACAA TCCAAAAATC ACAGGTTGG TGCACTGGCT	8040
	CACTCCTGTA ATCCCAACAC TTGGAAGGC CAAGGCAGGT GGATTGCTTC AGCTCAGGAG	8100
30	TTCAAGACCA GCCTGGGCAA CATGGCAAAA ACCCTGTCTC TCCAAAACAT ACACAAATTA	8160
	GCCAGATGTG GTAGTATGTG CCTGAGTCC CAACTACTCA AAAGGCTAAG GCAAGAGGAT	8220
	CACTTGAGCC CAGGAGGTCA AGGCTACAGT GAGCCATGTT TACTGTGTCA CTGCACTCCA	8280
	GCCTGGTCA TAGAGCAAGA CCATGCTCA AAAAAAAAGA AAAGAAAAGA AAAGAAAAAA	8340
	ACATCGCTCT ATTCAAGTCA CCCCCACAC ACATTTGTT TGATTATCAC ATAAATGCTG	8400
	GTCCATTGCC TTCTCTATCT ATTCAATCT TTAAGCATTC TTTGAGATTC AACTCAATT	8460
35	TCCTTTCAA ACTAGGCCAT TTAAACTACA TCAGTCCAT TTTGATTTTC TTGCTTTGAG	8520
	TCTACAGACT CAAAAACAAA AACTAAAAA CTTATTTTT AAGTTTCTG CTACTCTCAC	8580
	TTCTCAACA CTCACATACA CGCATTCTATA ATAAGATGGC AGAATGTTCA AGGATAAAAT	8640
	GATTTATAGA ACTGAAAAGT TAGGTTTGAA TCTTGTGCT GTCAAGATGA CTACCTACCT	8700
	GATCTCAGGT AATTAATTAT GTAGCATGCT CCCTCATTT ACCTACACC TATTCAACAG	8760
	GATTGGAATT CCACAGCAAG GATAAACACAT ATCATAGTT CTTTCAAGT TCAAGGCATT	8820
40	TTAACTTTA ATCTAGTACT ATGTTGTTG TTGTTGTTGT TGTTGAGAT GGAGCCCTGC	8880
	TGTGTACCC AGGCTGGAGT GCAGTGGCAC GAACTCGGGT CACTGCAACC TCTGCTCAT	8940
	GGGTTCAATC AGTTATTCTG CCTCAGTGTG CCAAGTAGCT GGGACTACAA GGCACATGCC	9000
	ACCATGCCGT GCTAATTTC GTATTTTAG TAGAACAGG GCTTCACCAT GTGGCCAGG	9060
	CTGGTCTCGA ACTCCTGACC TCAAGTGATC CAGCCGCCTC GGCCCTCCAA AGTGGCTGGGA	9120
45	TTACAGGCAT AAGCCACCGT GCCCCAGCTA ATAGTATGTT TTTAAACTCT TAGTGGCTTA	9180
	ACAATGCTGG TTGTATAATA AATATGCCAT AAATATTTAC TGTCTTAGAA TTATGAAGAA	9240
	GTGGTTACTA GGGCGTTTGC CACATATCAA TGGTTCTCTC CTTACAGCTT TAATTAGAGT	9300
	CTAGAATTGC AGGTTGGTAG AGCTGGAACA GACCTAAAG ATTGACTAGC CAACTCCTT	9360
	GTCCAATGA GGAAGCTGAG ACCCTAAAAA TTAAGTGACT TGCCCCAGAC AAAACTGGAA	9420
	CTCATGTGTC CTAATTCTCA TCATGAAATT CTACCATTC CAAGCCTCTG GCTAGTTGTC	9480
50	AAAGTATTGC ATAACATAAT TTTTATGTCT GTTTAAAGA ACAAAATTGTC ACTGCTTACT	9540
	CCTGGGAGGG TCTTCTGAG GTGGTTATA ACTCTAAAAA AAAAAAAAGT CAGTAGTCTG	9600
	AGAATTTTAG AGCAGGAAATG CAAAGCATTG TTATCCAATG GATCTATAAT TTTCATAGAT	9660

	TAGAGTTAAA	TCAAAGAAC	ACGGATGAGA	AAGGAAGAGG	AAAATGAGG	AGAGGAGGAA	9720
5	TGGGGATGAG	AACACACTAC	TTGTAATCG	TCATAGATGT	ACTGAGAACT	AACAAGAAGA	9780
	ATTGTAAGAA	AATAAGAATG	AAGAATTCAA	AATCAACACA	TGAAATAAAAA	ACAAACTACT	9840
	AGGGAAAAAT	GGAGAAGACA	TTAGAAAAAT	TATTCTATT	TTAAAATTCT	GTTTCAGGC	9900
	TTCCCTCCTG	TTCTTCCTCC	TTCTCATTGG	TTTCAGGTC	GAGGGAAAGT	TTAAGATGGA	9960
	AAAAATATAT	ATATTCTACA	CATCCCTTC	TACGCTGTT	TCATGCCAAC	AAGGTTTATC	10020
10	ATASCAAAC	TTTATTCTATA	CAACATTAT	TGAGTTCTTA	CTGTGTGGTA	AGCTCTTCC	10080
	ACGTGTTGAA	AATTCAAGGG	AAAAAAGACA	ACTCATTGTC	TTAAAACTCA	CATGAAAGCT	10140
	GAACGACCT	ATTTTTAATC	AAAGTAACT	CAATTAGGG	TAAGTAAAGAC	TATTTAAGAA	10200
	CCATGAACAG	GTGTGAAGGA	GGTAGGACTC	TGAGGAGAGA	ATAGTTAGCT	AGGAATGAAA	10260
	GAGCAGAGAA	GTTTTCTTAG	AGGAACATT	AAAGCTGGGA	GTACGGGAT	GAAGAGATGAG	10320
	GCAGGGTTG	CAGGCCAAA	AAAAAAAGG	GCAGGGGAAG	GGGAACCTCT	GCGCTGGCAG	10380
	AGAGAATAAC	TGTGGCAACA	ATGGAGGAGA	GTCTGGAAGC	AAGAAAACCA	AGTAAAGAG	10440
15	TATTTAAATA	GAAGATGCCA	GGGGTAATGA	GGGCTTGATT	TTAAACAGTG	CTGTTGGAGA	10500
	TCGAGAGGAG	ATACCAAATT	CTGGAGACAT	TTCTGAGTT	GAACCTACAG	TATTTATCAG	10560
	ACAAGGGAAA	GATTAGACAA	ACGAGTTAAG	AATGACTCCC	AGGTTCACT	TTGGCCGAGG	10620
	TAACTAGGAC	ATGTTTTGAA	AGTAATGTA	TTGGATCTCT	TACCATGGG	ACTATGTATG	10680
	TGGAGCCAA	TTAAAATTG	TACATGTATA	TAACCTCTCC	CCCACCCACCA	CTAACTACTT	10740
20	CCCTAACTCT	CTACTTTGTA	GCCAGACTTC	CTAAAAGAAAT	AGTTTGTAGT	CACTGTCTTT	10800
	ACTTTCCCC	TCCCATTCTG	TCCTAGATAT	TTGTCCACCT	ACCATCTGCT	CCCTCCACTT	10860
	TACCCAAACT	GTTCTACGGT	TGCCCCAAAC	TTCTTAATTG	CCAAATTCAA	TGAACAAAGT	10920
	TAAGCTTATA	TGAAATTAG	GAGCTCTACA	GTTCGATTTG	GAGCAGCCCC	TCCCTGAAACCC	10980
	CTTCTCTTT	CGACTTCTGT	GACACATCTC	AGATTTACAA	AACTGAACTA	ATTATTTTAC	11040
	ACTTGAGCTG	TATTTTCGT	CTTCTTCTT	GATGAATGAG	GTAACCACTC	AACAAATTGC	11100
	CCAAGCCAA	AACTACGAAG	TCATCCTCG	TTCCCTCTTC	TTCTGTTGAA	CCCACAAACAG	11160
25	ATCAGCTGAG	AAATCCCGCT	CTTGTAGTATC	TCTTGAATT	ATTACCTTAA	TTTATAGCCT	11220
	CATCAACTCT	TAATTGTTAA	AATTACTCA	GTACTTGTG	TCTGACCTCT	GTCCAATCTT	11280
	GTTCAATCG	GTCCATTCTT	TTGTTCTTGG	GGGTGGGGT	GGTGTGACA	GAGTTTCGCT	11340
	TTTGCTGCC	AGGCTGAAGT	GCAGTGGAGC	ACTTCACTGC	AACCACAGCC	TCCTGGGTT	11400
	AAGCAGTTCA	CCCTCCCGAG	TAGCTGGGAC	TACAGGTATG	TGCCACCCACA	CCCAGCTAAT	11460
	TTTGTGTTT	CAGTAGAGAC	AGGGTTTCAC	CATGTTGGTC	AGGCTGCTCT	CAAACCTCTC	11520
30	ACCTCAAGCA	ATCCACCCAC	CTCAGCCTCC	CAAAGTGCTG	CGATTACAGG	CATGAGCCAC	11580
	TGCACACGGA	CCAGATCCAT	TGTTTATGTT	GCTTCTAGAG	TGAGTTTTA	AAACACAAAT	11640
	TTGACCATAT	CTTCTCTCAA	TTAAAGTCAG	TATTTTTTT	TTCAGGAAAAA	AACAGTTCAA	11700
	ACTCTTCTAGT	CTGCTTACAC	AAGGCCCTTG	TAGTCTGACT	CTTCTTCCA	AGCTTTCATC	11760
	AAAGTATACT	GCAAGTTACA	TTTTATGTGA	ATTGAATTAG	GCAACGGTAT	AAAAATTATA	11820
	GTTTATATGG	GCAAAATGGA	ATAATGTTA	ACTCTTCCAA	ATAGTTTATC	TAGAATGACA	11880
35	TAATTCTAAA	GCTGTCAGGT	CAAATGAGTT	ATAAACTGTT	AACACTATTG	CCACATGCAA	11940
	GTGTCTCTTA	TACTTGGTAG	AATTATCTGC	TTCCATGTCA	TTATTATGTA	AATTAGACTT	12000
	TAAATAACTC	AGAAGTTCTT	CAGACATACA	GGTTATTATT	GTGCTTTTA	AACATAATT	12060
	TAAATAATT	TATATATGAT	AATGTATCC	AAAGTGCTAAG	GGATGTATTG	TTACTGCTGT	12120
	GCAAAAAAAA	AAAAAAAGAA	AACTCCAAT	AAATATGTTG	AAACCAAGTT	TATATGCAAG	12180
	AAAACAATAT	AAAAAAGGC	AAAGTACCCAC	CATAATAGGC	TGTGTTGGAGA	CGGCAGGCTA	12240
40	CAAAACACTA	GTAAATATGC	TGAGAAAGTT	AAAAAAAGAA	AGAAAGCAAC	AAATATGCTTT	12300
	GGTTGTTGTA	GGTTTATGTA	CTCCAAAGAAT	ATCTCCCTCT	AAACTTTAC	GTTTTTTCCA	12360
	AAGAAAAGT	AACTTTGGCT	GGGCGCAGTG	GCTCTTGCC	GTAGTCCCAG	CCTTTGGGAG	12420
	GCCAAGGCGG	GCAGATCACC	TGAGGTCAAGG	AGTTTGAGAC	CAGCCTGACC	AAAAATGGAG	12480
	AAACCCGCC	CCCTCACTAC	AAAAAGAATA	CAAATTAGG	CGGGCACAG	TGGCTTACCC	12540
45	CTGTGATCCC	AGCACTTTGG	GAGGCCGAAG	CAGGAAGATC	ACCTGAGGTC	AGGAGTTCGA	12600
	GACCAGCCAT	GGAGAAACCC	GTCTCTACTA	AAAATACAAA	ATTAGCCGGG	CGTGGTGGTG	12660
	CATGACTGTA	ATCCCAGCTA	CTCAGGAGGC	TAAGGCAGAG	AATCACTGTA	ACCCAGGGCAG	12720
	TGGAGGTTGC	AGTGAGGCCA	GATCGTGCCA	TTCCACTCCA	GCCTGGCAA	CAAGAGCGAA	12780
	ACTCTGTATC	AAAAAAACAA	AAGAAAAGAA	AAGGTAACCT	TGAACATATGT	GAGATCTTA	12840
	GAAATGCATT	CTTTCTGTAA	AATGTGACTA	CATTTGCCTT	ATTATGTTGA	AAAATGTTGA	12900
	GGCCTCAAAC	AACCCATATT	TTCTCGGTCT	CCCCGCTGCC	TAGCCTTTGT	TCACATTGCT	12960
50	TCTTCCTGGT	GGAAGCTCTT	CCTCTGGCCT	TGAAAATGCC	TGCTTCTCTT	TCAAGGTAGC	13020
	ACAGTCATCA	CTTTCTGTGG	TAACCTTCTC	CAGCACCATC	AAACAGAAAG	AATGAATCTC	13080

	TTGTAAATTC	AGCTCTTACG	TCATTCCATTA	CATTATTTG	TAACCTTTA	TAGATTCTTC	13140
5	TCTCCCACTA	GAATCTGAGT	CACTGGAGAG	TAGGAGCCAA	CTCTCATTCA	TGTGTGGTTT	13200
	GGTCAGCTAC	TGGCCACATT	CCTGATGCAT	AGTTAATGCT	CAAACCTTAA	CTGGTGAATC	13260
	AGCTCAAATA	TTGTCCCTTCT	CTAAATCCAT	TCACTCATG	ACTAACTATG	TACTCAAAT	13320
	AGTAAACACC	AGTAATTAA	TCCAATTCCCT	CCCCATACCTC	CTTCGTACAT	ITCAGGTGAA	13380
	TTAGTTGAT	AAATATGTGT	GTATTACATA	ATATTAAAGT	ATGTACAGAA	GATCATGCTA	13440
	ATCATTAATTC	ACAACGTATA	ACTAACCAA	CATAAATGCT	CTCAGGTTAA	CAAATGTCTG	13500
10	CCTTCTCAGT	TAATGCAGTC	ATTAACAAAC	ACCTTCTGAT	GCTGATAATA	GGGCCTTGT	13560
	CAGCAATGAA	CCCATAAAGG	TGAATAAAGA	ACATGCCCTC	GTGGAGCTCA	CAGCCTAGTC	13620
	ATTATTGTTTC	TGATTTTTAA	TATTAATGTT	GGTTTGGTT	TTGGTGAAAAA	ATGTTTAGAC	13680
	TTATCTTAGT	GATCTTTCA	TCCTTGCTA	TATTATTTT	CTCTAAGAGT	CTTCCTTATC	13740
	CCCTCCTTAA	AAAAACTAGG	TGATAATTCT	AAATTGTAAA	TTTAAATATT	ATAAAATAGCT	13800
	TATAAAATT	AAATATTATA	ATATTAAAT	GTGGATAAA	TATTTAAATT	TTATAATATT	13860
15	TAAATGTTTA	TTTAAATTCA	TTTGTACATC	AGTTTTTATT	TTATTTAAAT	GTGTTGGCCA	13920
	GGCATGGTGG	CTGACACCTA	TAATCCCAGA	ACTTTGAGAG	GCCAAGTCAG	GCAAACCATT	13980
	TGAGCTCAGG	AGTTTGAGAC	CACCCCTGGGC	AACGTGGTGA	AACCCGTCT	CTACCAAACA	14040
	TATGAAAAC	TATCTGGGTG	TGGTGGCACG	CATCTGTGGT	CCCAGATGGG	AGTCCCAGGC	14100
	TAAGATGGGA	GAATCGCTTG	AACCCAGCTG	AGAGGGCTGG	GGTGGATGTT	GCAGTGAGCT	14160
	GAGATCGTC	CACTGCACTC	CAACCTGGGT	GACAGASTGA	GACTCCATCT	CAAAAAA	14220
20	AAATGTTATC	AAATAAAGAT	AAATTAAATA	ACTGTTCGA	CTTAGATGAG	CATAAGGAAC	14280
	TAAACCTAGA	AAAACATATC	AAAATAAGGCC	TGGGTACAGT	GACTCATGCC	TGTAATCTCA	14340
	AGCACCTTGG	GAGGCCAAA	TTATACAAAG	TTAGTTGTAT	AACACCAACT	ACAACATT	14400
	TTGGGGTTAG	CTTAATTCA	ATTAATTTTT	TTTAAACTGTA	GTGTTAAATT	CCTGCTTACT	14460
	CTACCATACA	TGCTAGGCCT	CATATTATGC	TAGAAAAATT	TTGACACAG	ATTTATGAAT	14520
	ACTCTCCTGC	ATACCATTAA	ATTTTAAAC	AAATTAAAT	GCAGTATATA	TGTGCCCTT	14580
25	TACCAACACA	TTAAATAATA	AGATCTACTG	TGAGGACTAA	ATTTCTGTAA	TTTCAAAGTA	14640
	GTAAATGAGTT	AAACCACATGT	CTCAAGATCT	CTGCAATAAC	TGTAGCACAA	CAGAAAATAG	14700
	GTATTCTCAT	TAATGACAGA	GTCACAAAGTA	CTACTAATAA	TACTGTGGTT	TGTTTCCTGC	14760
	AACTAATCAT	GGGAGGAATG	CTAAATTCA	GAGGTTGGTG	AAAATACATG	TGTATTTTT	14820
	TCCCCATCCA	AGTTCACAGA	TTCTCACAC	TGAGAACTCC	TATTCATCAA	CAAAATTCTG	14880
	GAAGCCTGCA	CACCGTATTG	GAAGAAGGGG	AGAAAGGAAA	AGCAAATGGA	AGGATTTAA	14940
30	TTTTTTCAA	ATCCTGTATC	CCTTGATTTT	ACACCAAGAT	TGTATTTATG	TATTACTTGT	15000
	GTAAAAAATA	TAGTATAATC	GAGACTCCAG	ATCAAAATC	ACCGCAGCTC	AGGGAGAAAG	15060
	AGGGCCACCA	AATGCCAGAG	CCCTTCAGCC	TTCTCCCACC	CTGCCTGTAC	CCTCAGATGG	15120
	AAGCACTTTT	TTATCATTGT	TTCACCTTTA	GCATTTTGAC	AATGAAGTCA	CAAACCTTCA	15180
	GCCTCTCACC	CATAGGAACC	CACTGGTTGT	AAGAGAAGGA	TGAAGCCAGT	CCTTCCTAA	15240
	GGGCACGATT	AGATGTGTTT	ATGGCATTCT	CAGGTGAAAC	TATATTATA	TTGACAATAT	15300
35	ATTATATTT	CTCAAGGAAT	ACTAGAAATAA	TGATTCAAGT	CAGTACTAGG	CCATTTATCT	15360
	ACCCCTTATA	ATATTGTTA	ATGAGAAAAT	GCTTCTATC	TTCCAAATAT	CTGATGATT	15420
	CTAAGAGAAC	ACTAAACAT	GGGTATTCA	AAGCTGAAAC	TTCTGGCATT	TATTGAATGT	15480
	CAAGATTGTT	CATCACTA	CTAGGTGATT	AACTGACCAC	TGAACCTTGAA	GGTAGTATAA	15540
	AGTAGTAGTA	AAAGGTACAA	TCATTGTCTC	TTAACAGATG	GCTCTTGT	TTCATTAGGA	15600
40	ATAAAG ATG	GCT GCT GAA	CCA GTA GAA GAC	AAT TGC ATC AAC	TTT GTG GCA	15651	
	Met Ala Ala Glu	Pro Val Glu	Asp Asn Cys	Ile Asn Phe	Val Ala		
	-35	-30	-25				
	ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G			CTAAGGC	TAATGCCATA	15702	
	Met Lys Phe Ile Asp Asn Thr Leu Tyr	Phe Ile Ala					
	-20	-15	-10				
45	GAACAAATAC	CAGGTTCAGA	TAAATCTATT	CAATTAGAAA	AGATGTTGTG	AGGTGAACTA	15762
	TTAAGTGA	CTTTGTGTCA	CCAAATTCA	CTGTAATATT	AATGCCCT	AAAAAAATAG	15822
	TGGACCTCTA	GAAATTAACC	ACAACATGTC	CAAGGTCTCA	GCACCTTGTC	ACACCACGTG	15882
	TCCTGCACT	T'ATATCAGCA	GTAGCTCACT	CTCCAGTTGG	CAGTAAGTGC	ACATCATGAA	15942
	AATCCCAGTT	TTCATGGGAA	AATCCAGTT	TTCATTGGAT	TTCCATGGGA	AAAATCCAG	16002
	TACAAAATCT	GGTGCATTCA	GGAAATACAA	TTTCCCAAAG	CAAATGGCA	AATTATGTAA	16062
50	GAGATTCTCT	AAATTAGAG	TTCCGTGAAT	TACACCATT	TATGAAATA	TGTTTGACAA	16122
	GTAAAAATTG	ATTCTTTTTT	TTTTTTCTG	TTGCCAGGGC	TGGAGTGCAG	TGCCACAAATC	16182
	TCTGCTCACT	GCAACCTCCA	CCTCTGGGT	TCAACCAATT	CTCCTGCCCT	AGCCTTCTGA	16242

5	GTAGCTGGGA	CTACAGGTGC	ATCCCCCAT	GCCTGGCTAA	TTTTTGGGTA	TTTTTACTAG	16302
	AGACAGGGTT	TTGGCATGTT	GTCCAGGCTG	GTCTTGGACT	CCTGATCTCA	GATGATCCTC	16362
	CTGGCTCGGG	CTCCCAAAGT	GCTGGGATTA	CAGGCATGAA	CCACCACACA	TGGCCTAAAAA	16422
	ATTGATTCTT	ATGATTAATC	TCCTGTGAAC	AATTTGGCTT	CATTGAAAG	TTTGCCTTCA	16432
	TTTGAAACCT	TCATTTAAAAA	GCCTGAGCAA	CAAAGTGAGA	CCCCATCTCT	ACAAAAAAACT	16542
	GCAAAATATC	CTGTGGACAC	CTCCTACCTT	CTGTGGAGGC	TGAAGCAGGA	GGATCACTTG	16602
	AGCCTAGGAA	TTTGAGCCTG	CACTGAGCTA	TGATCCCACC	CCTACACTCC	AGCCTGCATG	16662
	ACAGTAGACC	CTGACACACCA	CACACAAAAA	AAAACCTCTCA	TAACAAATTAA	TTAGTTGACT	16722
10	TTTCTTAGGT	GACTTTCCGT	TTAAGCAATA	AATTTAAAAG	AAAATCTCT	AATTTTAGAA	16782
	AATTTATTTT	TAGTTACATA	TTGAAATTTT	AAACCCCTAG	GTAAAGTTT	TATGTCTAAA	16842
	TTACCTGAGA	ACACACTAAG	TCTGATAAGC	TTCAATTCTAT	GGGCCTTTTG	GATGATTATA	16902
	TAATATTCTG	ATGAAAGCCA	AGACAGACCC	TTAAACCATA	AAAATAGGAG	TTCGAGAAAG	16962
	AGGAGTAGCA	AAACTAAAAG	CTAGAATGAG	ATTCAATTCT	GACTCGAAAT	ACAAAATTTT	17022
	ACATATTCTC	TTTCTCTCTT	TTAG	CT CAA GAT GAT G	GTAAACT		17075
15					Ala Glu Asp Asp Glu		
					-10		
	AGAAATGAAT	TTATTTTCT	TTGCAAACTA	AGTATCTGCT	TGAGACACAT	CTATCTCACC	17135
	ATTGTCAGCT	GAGGAAAAAA	AAAATGGTT	CTCATGCTAC	CAATCTGCCT	TCAAAGAAAT	17195
	CTGCACTCAG	TACCAACAGCT	TTGGAATGAA	GATGATCATA	AGAGATACAA	AGAAGAACCT	17255
20	CTAGCAAAAG	ATGCTTCTCT	ATGCCCTAAA	AAATTCTCCA	GCTCTAGAA	TCTACAAAAT	17315
	AGACTTTGCC	TGTTTCATTG	GTCTAACAGT	TAGCATGAAG	CCATGGATTC	TGTTGTAGGG	17375
	GGAGCGTTGC	ATAGGAAAAA	GGGATTGAAG	CATTAGAATT	GTCCAAATC	AGTAACACCT	17435
	CCTCTCAGAA	ATGCTTGGG	AAGAACCTG	GAAGGTTCCG	GGTTGGTGGT	GGGGTGGGGC	17495
	AGAAAATCT	GGAAGTAGAG	GAGATAGGAA	TGGGTGGGGC	AAGAAGACCA	CATTCAAGAGG	17555
	CCAAAACCTG	AAAGAAACCA	TGGCATTTTAT	GATGAATTCA	GGGTAAATTCA	GAATGGAAGT	17615
25	AGAGTAGGAC	TAGGAGACTG	GTGAGAGGGAG	CTAGAGTGT	AAACAGGGTG	TAGAGCAAGA	17675
	CGTTCTCTCA	CCCCAAGATG	TGAAATTG	ACTTTATCTT	GGAGATAA	GGGTTAATT	17735
	AGCACAAATAT	GTATTAGCTA	GGGTAAAGAT	TAGTTTGTG	TAACAAAGAC	ATCCAAAGAT	17795
	ACAGTAGCTG	AATAAGATAG	AGAATTCTTC	TCTCAAAAGA	AGTCTAAGT	GGCAGCTCAG	17855
	AAGTACTATG	GCTGGAAGCA	ACCTGTGATG	ATTGGGACCC	CCAACCTCT	TCAGTCTTGT	17915
	ACCCATCATC	CCCTAGTTGT	TGATCTCACT	CACATAGTTG	AAAATCATCA	TACTTCCTGG	17975
30	GTTCATATCC	CAGTTATCAA	GAAAGGGTCA	AGAGAAGTC	GGCTCATTCC	TTTCAAAGAC	18035
	TCTAATTGGA	AGTTAAACAC	ATCAATCCCC	CTCATATTCC	ATTGACTAGA	ATTTAATCAC	18095
	ATGGCCACAC	CAAGTGCAG	GAAATCTGGA	AAATATAATC	TTTATTCCAG	GTAGCCATAT	18155
	GACTCTTAA	AATTCAAGAAA	TAATATATT	TTAAAATATC	ATTCTGGCTT	TGGTATAAAAG	18215
	AATTGATGTT	GTGGGGTGTAG	GAGGCCAAA	TTAAGGTTG	AGAGCCTATT	ATTTTAGTTA	18275
	TTACAAGAAA	TGATGGTGT	ATGAATTAAG	GTAGACATAG	GGGAGTGCTG	ATGAGGAGCT	18335
35	GTGAATCGAT	TTAGAAACAC	CTTGAGAGAA	TCAATAGGAC	ATGATTTAGG	GTTGGATTG	18395
	GAAAGGAGAA	GAAACTAGAGA	AAGATGATGC	CTACATTTTT	CACTTAGGCA	ATTTGTACCA	19455
	TTCAGTGAAGA	TAGGGAACAC	AGGAGGAAGA	GCAGGTTTG	GTCTATACAA	AGAGGAGGAT	18515
	GGATGACGCA	TTTCGTTTG	GATCTGAGAT	GTCTGTGGAA	CGTCCTAGTG	GAGATGTCCA	18575
	CAAACCTCTC	TACATGTGGT	TCTGAGTTCA	GGACACAGAT	TTGGGTGGA	GATAGAGATA	18635
	TTGTAGGCTT	ATACATAGAA	ATGGCATTG	AATCTATAGA	GATAAAAAGA	CACATCAGAG	18695
40	GAAATGTGTA	AAAGTGAAGAGA	GGAAAAGCCA	AGTACTGTGC	TGGGGGAAT	ACCTACATT	18755
	AAAGGATGCA	GTAGAAAGAA	GCTAATAAAC	AAACAGAGAC	AGACTAACCA	AAAGGGGAGA	18815
	AGAAAACCA	AGAGAATTCC	ACCGACTCCC	AGGAGAGCAT	TTCAAGATTG	AGGGGATAGG	18875
	TGTTGTGTTG	AATTTTGCA	CCTTGAGAA	CAAGGGCCAG	AACACAGCTT	TTAGATTTAG	18935
	CAACAAGGAG	TTGGTGATC	TCAGTGAAG	CAGCTTGATG	GTGAAATGGA	GGCAGAGGCA	18995
	GATTGCAATG	AGTGAAACAG	TGAATGGAA	GTGAAGAAAT	GATACAGATA	ATTCTTGCTA	19055
45	AAAGCTTGGC	TGTTAAAAGG	AGGAGAGAAA	CAAGACTAGC	TGCAAAGTGA	GATTGGGTTG	19115
	ATGGGACCACT	TTAAATCTC	AAAATAAAGA	GCTTTGTGCT	TTTTGATTA	TGAAAATAAT	19175
	GTGTTAATTG	TAACTAATTG	AGGCAATGAA	AAAAGATAAT	AATATGAAAG	ATAAAAATAT	19235
	AAAAACCAAC	CAGAAATAAT	GATAGCTACC	ATTTGATAC	AATATTCTCA	CACTCCTTTC	19295
	TATGTATATA	TACAGACACA	GAAATGCTTA	TATTTTTAT	AAAAGGGATT	CTACTATACC	19355
	TAAGCTGCTT	TTCTCTAGTTA	GTGATATATA	TGGACATCTC	TCCATGGCAA	CGAGTAATTG	19415
50	CAGTTATATT	AAGTTCATGA	TATTCACAA	TAAGGGCATA	TCTTTGCCCT	TTTTTATTTAA	19475
	TCAATTCTTA	ATTGGTGAAT	GTGTTTTCC	AGTTTGTGTTG	TGTTATTAAC	AATGTTCCCA	19535

	TAAGCATTCC	TGTACACCAA	TGTTCACACA	TTTGTCTGAT	TTTTTCTTCA	GGATAAAACC	19595
	CAGGAGGTAG	AATTGCTGGG	TTGATAGAAG	AGAAAGGATG	ATTGCCAAT	TAAAGCTCA	19655
5	GTAGAGGGTA	CATGCCGAGC	ACAAATGGGA	TCAGCCCTAG	ATACCAGAA	TGGCACTTTC	19715
	TCATTTCccc	TTGGGACAAA	AGGGAGAGAG	GCAATAACTG	TGCTGCCAGA	GTAAATTG	19775
	TACGTGGACT	AGCAGGAAT	CATTGCTGA	AAATGAAAC	AGAGATGATG	TTGTAGAGGT	19835
	CCTGAAGAGA	GCAAAGAAA	TTTGAATTG	CGGCTATCAG	CTATGGAAGA	GAGTGCCTGAA	19895
	CTGGAAAACA	AAAGAACGTAT	TGACAATTG	TATGCTTGT	ATGCCACCAGA	TTTGAACGCT	19955
10	TGTGCCATTG	TTCACCAGCA	GCACTCAGCA	GCCAAGTTG	GAGTTTGT	GCAGAAAGAC	20015
	AAATAAGTTA	GGGATTTAAT	ATCCTGGCCA	AATGGTAGAC	AAAATGAAC	CTGAGATCCA	20075
	GCTGCACAGG	GAAGGAAGGG	AAGACGGGA	GAGGTAGAT	AGGAAATACA	ACAGTCAGGA	20135
	GAATGGAAGA	TGTTGTGATA	TTAAGAACAA	CATAGAGTTG	GAGTAAAAGT	GTAAGAAAAC	20195
	TAGAAGGGTA	AGAGACCGGT	CAGAAAGTAG	GCTATTGAA	GTAAACACTT	CAGAGCCAGA	20255
	GTAGTTCTGA	ATGGTAACAA	GAAATTGAGT	GTGCCTTGT	GAGTAGGT	AAAAACAAATA	20315
	GGCAACTTTA	TTGTAGCTAC	TTCTGGAACA	GAAGATTGTC	ATTAATAGTT	TTAGAAAAC	20375
15	AAAATATATA	GCATACTTAT	TTGTCATTAA	ACAAAGAAC	TATGTATTT	TAAATGAGAT	20435
	TTAATGTTA	TTGTAG	AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA				20486
			Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu				
			-5	1	5		
	TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT						20534
	Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile						
20	10	15	20				
	GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT						20582
	Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys						
	25	30	35				
	AGA G GT ATTTTTTTA ATTCCAAAC ATAGAAATGA CTAGCTACTT CTTCCCATTC						20638
25	Arg Asp						
	40						
	TGTTTACTG CTTACATTGT TCCGTCTAG TCCCAATCCT CAGATGAAA GTCACACGGAG						20698
	TGACAATAAT TTCACTTACA GGAAACTTTA TAAGGCATCC ACCTTTTTA GTTGGGGTAA						20758
	AAAATTGGAT ACAATAAGAC ATTGCTAGGG GTCATGCCCTC TCTGAGCCTG CCTTTGAATC						20818
30	ACCAATCCCT TTATTGTGAT TGCATTAACCT GTTTAAACCC TCTATAGTTG GATGCTTAAAT						20878
	CCCTGCTTGT TACAGCTGAA AATGCTGATA GTTTACCAAG GTGTTGGCA TCTATCTGTA						20938
	ATCCTAGCTA CTTGGGAGGC TCAACCGAGGA GGATTGCTTG AGGCCAGGAC TTTGAGGCTG						20998
	TAGTACACTG TGATCGTACC TGTGAATAGC CACTGCACTC CAGCCTGGGT GATATACAGA						21058
	CCTTGTCTCT AAAATTAAAAA AAAAAAAA AAAAACCTT AGGAAAGGAA ATTGATCAAG						21118
	TCTACTGTGC CTTCCAAAC ATGAATTCCA AATATCAAAG TTAGGCTGAG TTGAAGCAGT						21178
	GAATGTGCAT TCTTTAAAAA TACTGAATAC TTACCTTAAC ATATATTAA AATATTAAAT						21238
35	TTAGCATTTA AAAGTTAAAAA ACAATTTTT AGAATTCTATA TCTTTAAAT ACTCAAAAAA						21298
	GTTGCAGCGT GTGTGTTGTA ATACACATTA AACTGTGGGG TTGTTTGTGTT GTTGAGATG						21358
	CAGTTTCACT CTGTCACCCA GGCTGAAGTG CAGTGCAGTG CAGTGGTGTG ATCTCGGCTC						21418
	ACTACAACCT CCACCTCCCA CGTTCAAGCG ATTCTCATGC CTCAGTCTCC CGAGTAGGTG						21478
	GGATTACAGG CATGCCAAC TTACACCCGG CTAATTGTTG TATTGTTAGT AGAGCTGGGG						21538
40	TTTCACCATG TTGGCCAGGC TGGTCTCAAA CCCCTAAACT CAAGTGTACT GCCTGCCTCA						21598
	GCCTCCAAA CAAACAAACA ACCCCACAGT TTAATATGTG TTACACACAA CATGCTGCAA						21658
	CTTTTATGAG TATTTAAATG ATATAGATA TAAAAGGTG TTTTTAAACTT TTAAATGCTG						21718
	GGATTACAGG CATGAGGCCAC TGTGCCAGGC CTGAACCTGTG TTTTTAAAGA TGTCTGACCA						21778
	GCTGTACATA GTCTCCTGCA GACTGGCCAA GTCTCAAAGT GGGAACAGGT GTATTAAGGA						21838
	CTATCCTTTG GTAAATTTC CGCAATGTT CCTGTGCAAG AATTCTTCTA ACTAGAGTT						21898
45	TCATTTATTA TATTTATTTC AG AT AAT GCA CCC CGG ACC ATA TTT ATT ATA						21949
	Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile						
	40	45					
	AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT						21997
	Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile Ser						
50	50	55	60				22045
	GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT						
	Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile						
	70	75	80				

	TCC TTT AAG GTAAGACTG AGCCTTACTT TGTTTCAAT CATGTTAATA TAATCAATAT	22103
	Ser Phe Lys	
5	AATTAGAAAT ATAACATTAT TTCTAAATGTT AATATAAGTA ATGTAATTAG AAAACTCAAA	22163
	TATCCTCAGA CCAACCTTT GTCTAGAACAA GAAATAACAA GAAGCAGAGA ACCATTAAAG	22223
	TCAATACTTA CTAAAAATTAA TCAAACCTT TACCTATTGT CATAATGATG GTTTTCTGA	22283
	GCCTOTCACA CGCGAAGAGG AGATACAAACAA CTTGTTTAT GACCTGCATC TCCTCAACAA	22323
	TCAGTCTTA TACAAATAAT AATGTAGAAT ACATATGTGA GTTATACATT TAAGAATAAC	22403
	ATGTGACTTT CCAGAATGAG TTCTGCTATG AAGAATGAAG CTAATTATCC TTCTATATT	22463
10	CTACACCTT GTAAATTATG ATAATATTTT AATCCCTAGT TGTTTGTG CTGATCCTTA	22523
	GCCTAAGTCT TAGACACAAG CTTCAGCTC CAGTTGATGT ATGTTATTTT TAATGTTAAT	22583
	CTAATTGAAAT AAAAGTTATG AGATCAGCTG TAAAAGTAAT GCTATAATTA TCTTCAAGCC	22643
	AGGTATAAAAG TATTTCTGGC CTCTACTTT TCTCTATTAT TCTCCATTAT TATTCTCTAT	22703
	TATTTTCTC TATTTCTCC ATTATTGTTA GATAAACCAC AATTAACATAT AGCTACAGAC	22763
15	TGAGCCAGTA AGAGTAGCCA CGGATGCTTA CAAATTGGCA ATGCTTCAGA GGAGAATTCC	22823
	ATGTCATGAA GACTCTTTT GAGTGGAGAT TTGCCAATAA ATATCCGCTT TCAAGCCCCAC	22893
	CCAGTCCCCA CTGAAAGACA GTTAGGATAT GACCTTAGTG AAGGTACCAA GGGGCAACTT	22943
	GGTAGGGAGA AAAAAGCCAC TCTAAATAT AATCCAAGTA AGAACAGTGC ATATGCAACA	23003
	GATACAGCCC CCAGACAAAT CCCTCAGCTA TCTCCCTCCA ACCAGAGTGC CACCCCTTCA	23053
	GGTGACAATT TGGAGTCCCC ATTCTAGACC TGACAGGCAG CTTAGTTATC AAAATAGCAT	23123
20	AAGAGCCCTG GGATGGAAGG GTAGGGTGG AAGGGTTAAC CATGCTGTTA CTGAACAAACA	23183
	TAATTAGAAG GGAAGGGAGAT GGCCAAGCTC AAGCTATGTC GGATAGAGGA AAACCTAGCT	23243
	CCAGAGGCAG ATTCAAGAAC TGGGATAACT CCGAACCTAC AGGTGGATTC TTGTTGAGGG	23303
	AGACTCGTGA AAATGTTAAC AAGATGGAAA TAATGCTTGG CACTTACTAC GAACTGGCA	23363
	AATCCATATT TGGGGGAGCC TGAAGTTAT TCAATTITGA TGGCCCTTTT AAATAAAAAG	23423
	AATGTGGCTG GCGTGTGGTGG CTCACACCTG TAATCCCAGC ACTTTGGAG GCCGAGGGGG	23483
25	GCGGATCACC TGAAGTCAGG AGTTCAAGAC CAGCCTGACC AACATGGAGA AACCCCATCT	23543
	CTACTAAAAA TACAAAATTAA GCTGGCGTG GTGGCATATG CCTGTAATCC CAGCTACTCG	23603
	GGAGGCTGAG GCAGGAGAAAT CTTTCAACC CGGGAGGCAG AGGTTGCGAT GAGCCTAGAT	23653
	CGTGCCTATTG CACTCCAGCC TGGGCAACAA GAGCAAAACT CGGTCTCAA AAAAAAAA	23723
	AAAAAGTGA ATTAAACAAA GGCATTAGCT TAATAATTAA ATACTGTTT TAAGTAGGGC	23783
	GGGGGGTGGC TCGAAGAGAT CTGTGTAAT GAGGGAAATCT GACATTAAAG CTTCATCAGC	23843
30	ATCATAGCAA ATCTGCTTCT CGAAGGAACT CAATAAATAT TAGTTGGAGG GGGGGAGAGA	23903
	GTGAGGGGTG GACTAGGACC AGTTTAGCC CTTGTCCTTA ATCCCTTTTC CTGCCACTAA	23963
	TAAGGATCTT AGCAGTGGTT ATAAAAGTGG CCTAGGTTCT AGATAATAAG ATACAACAGG	24023
	CCAGGCACAG TGGCTCATGC CTATAATCCC AGCACTTTGG GAGGGCAAGG CGAGTGTCTC	24083
	ACTTGAGATC AGGAGTTCAA GACCAGCTG GCCAGCATGG CGATACTCTG TCTCTACTAA	24143
	AAAAAATACA AAAATTAGCC AGGCATGGTG GCATGCACCT GTAATCCCAG CTACTCGTGA	24203
35	GCCTGAGGCA GAAGAATCGC TTGAAACCAG GAGGTGTAGG CTGCAGTGAG CTGAGATCGC	24263
	ACCACTGCAC TCCAGCCTGG GCGACAGAAAT GAGACTTTGT CTAAAAAAA GAAAAGATA	24323
	CAACAGGCTA CCCCCTATGTG CTCACCTTTC ACTGTTGATT ACTAGCTATA AAGTCTCTATA	24383
	AAGTTCTTTG GTCAAGAACAC TTGACACAC TAAGGGGGAT TTGCTTGAG AGTTACTGT	24443
	CAGAGTCTGT TTCATATATA TACATATACA TGTATATATG TATCTATATC CAGGCTTGGC	24503
	CAGGGTTCCC TCAGACTTTG CAGTGCACCTT GGGAGATGTT AGGTCAATAT CAACTTCCC	24563
40	TGGATTCAAGA TTCAACCCCT TCTGATGTA AAAAAAAA AAAAAAGAAA GAAATCCCTT	24623
	TCCCCTTGGG GCACTCAAGT TTCACCAAGGT GGGGCTTCC AAGTTGGGG TTCTCCAAGG	24683
	TCATTGGGAT TGCTTTCACA TCCATTGCT ATGTACCTTC CCTATGATGG CTGGGAGTGG	24743
	TCAACATCAA AACTAGGAAA GCTACTGCCA AAGGATGTCC TTACCTCTAT TCTGAAATGT	24803
	GCAATAAGTG TGATTAAGA GATTGCCTGT TCTACCTATC CACACTCTCG CTTCAACTC	24863
	TAACTTTCTT TTTTTCTTTT TTTCTTTTTT TGAAACGGAG TCTCGCTCTG	24923
45	TCGCCAGGC TAGAGTGCAG TGGCAGGATC TCAGCTCACT GCAAGCTCTG CCTCCCGGGT	24983
	TCACGCCATT CTCCTGCCTC ACCCTCCAA GCAGCTGGGA CTACAGGCAC CTGCCACCAT	25043
	GCCCAGCTAA TTTTTGTAT TTTTAGTGA GACGGGGTTT CACCGTGTAA GCCAGGATGG	25103
	TCTCGATCTC CTGAACTTGT GATCCCCCG CCTCAGCCTC CCAAAGTGCT GGGATTACAG	25163
	GGGTGAGCCA TCGCACCCGG CTCAACTGTA ACTTTCTATA CTGGTTCATC TTCCCTGTAA	25223
50	ATGTTACTAG AGCTTTGAA GTTTGGCTA TGATTATTT CTCATTATA CATTAGATTT	25283
	CAGATTAGTT CCAAATTGAT GCCCACAGCT TAGGGTCTCT TCCTAAATTG TATATTGTAG	25343
	ACAGCTGCAG AAGTGGGTGC CAATAGGGGA ACTAGTTAT ACCTTCATCA ACCTAGGACC	25403

	CACACTTGTT GATAAAGAAC AAAGGTCAAG AGTTATCACT ACTGATTCCA CAACTGATTG 25463
	AGAACTTGGG AATAACCCCG TGACCTCTGC CATCCAGAGT CTTTCAGGCA TCTTTGAAGG 25523
5	ATGAAGAAAT GCTATTTAA TTTGGAGGT TTCTCTATCA GTGCTTAGGA TCATCGGAAT 25583
	CTGTGCTGCC ATGAGGCCAA AATTAAGTCC AAAACATCTA CTGGTCCAG GATTAACATG 25643
	GAAGAACCTT AGGTGGTGGC CACATGTTCT GATCCATCCT GCAAAATAGA CATGCTGCCAC 25703
	TAACAGAAA AGTCAGGCCA GCACTACCAG TTGGATAACC TGCAAGATTA TAGTTCAAG 25763
	TAATCTAACCTTCTACA AGGCCTTATT CTGTGACTGA AACATACAAG AATCTGCATT 25823
	TGGCCTCTA AGGCAGGGCC CAGCCAAGGA GACCATATTC AGGACAGAAA TTCAAGACTA 25883
10	CTATGGAACCTT GGAGTGCTTG GCAGGGAAAG CAGAGTCAAC GACTGCCAAC TGAGCCATA 25943
	CAGCAGGCTT ACACAGGAAC CCAGGGCTA CCCTACAAAC AATTATTGGG TCTATTCACT 26003
	GTAAGTTTA ATTTCAGGCT CCACTGAAAG AGTAAGCTAA GATTCCTGGC ACTTTCTGTC 26063
	TCTCTCACAG TTGGCTCAGA AATGAGAACT GGTCAAGGCCA GGCATGGTGG CTTACACCTG 26123
	GAATCCCAGC ACTTTGGGAG GCCGAAGTGG GAGGGTCACT TGAGGCCAGG AGTCAGGAC 26183
15	CAGCTTAGGC AACAAAGTGA GATACCCCT GACCCCTTCT CTACAAAAAT AAATTTAAA 26243
	AATTAGCCAA ATGTGGTGGT GTATACTTAC AGTCCCAGCT AACTCAGGAGG CTGAGGCAGG 26303
	GGGATTGCTT GAGCCCAGGA ATTCAAGGCT GCAGTGAGCT ATGATTTCAC CACTGCACTT 26363
	CTGGCTGGC AACAGAGCCA GACCCCTGTCT CAAAGCAAAA AGAAAAAGAA ACTAGAACTA 26423
	GCCTAAGTTT GTGGGAGGAG GTCATCATCG TCTTAGCCG TGAATGGTTA TTATAGAGGA 26483
	CAGAAATTGA CATTAGGCCA AAAAGCTTGT GGTCTTGCT GGAACCTCAC TTAATCTTGA 26543
	GCAAATGTGG ACACCACTCA ATGGGAGAGG AGAGAAGTAA GCTGTTGAT GTATAGGGGA 26603
20	AAACTAGAGG CCTGGAACCTG AATATGCATC CCATGACAGG GAGAATAGGA GATTGGAGT 26663
	TAAGAAGGAG AGGAGGTCAAG TACTGCTGTT CAGAGATTAA TTTTATGTAA CTCTTGAGAA 26723
	GCAAAACTAC TTTTGTCTG TTTGTAATA TACTTCAAA CAAACTTCAT ATATTCAAAT 26783
	TGTCATGTC CTGAAATAAT TAGGTAATGT TTTTCTCT ATAG GAA ATG AAT CCT 26839
	Glu Met Asn Pro
	85
25	CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA 26897
	Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg
	90 95 100
	AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC 26935
	Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Tyr
	105 110 115 120
30	GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC 26983
	Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu
	125 130 135
	ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT 27031
	Ile Leu Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr
	140 145 150
35	GTT CAA AAC GAA GAC T AGCTATTAAA ATTCATGCC GGGCGCAGTG GCTCACGCCT 27087
	Val Gln Asn Glu Asp
	155
	GTAATCCCAG CCCTTGGGA GGCTGAGGCG GGCAGATCAC CAGAGTCAG GTGTTCAAGA 27147
	CCAGCCTGAC CAACATGGTG AAACCTCATC TCTACTAAAA ATACAAAAAA TAGCTGAGT 27207
40	GTAGTGACCC ATGCCCTCAA TCCCAGCTAC TCAAGAGGCT GAGGCAGGAG ATCACTTGC 27267
	ACTCCGGAGG TGGAGGTTGT GGTGAGCCGA GATTGACCA TTGCGCTCTA GCCTGGCAA 27327
	CAACAGCAAA ACTCCATCTC AAAAATAAA ATAAATAAA AAACAAATAA AAAATTCTATA 27387
	ATGTGAAC TCTGAATTAA TATGTTAGA AAGATTATGA GATTATTAGT CTATAATTGT 27447
	AATGGTGGAA TAAAATAAA ACCAGTCTTG AAAAACATCA TTAAGAAATG AATGAACCTT 27507
	CACAAAAGCA AACAAACAGA CTTCCCTTA TTTAAGTGA TAAAATAAA TAAAATAAA 27567
45	TAATGTTAA AAAATTCTATA GTTTGAAAC ATTCTACATT GTTAATTGGC ATATTAATTA 27627
	TACTTAATAT AATTATTTT AAATCTTTG GGTTATTAGT CCTAATGACA AAAGATATTG 27687
	ATATTGAAAC TTCTAATT TTAAGAATAT CGTTAAACCA TCAATATTAA TATAAGGAGG 27747
	CCACTTCACT TGACAAATT CTGAAATTCC TCCAAAGTCAGTATTTT AAAATTCACT 27807
	TTGATCCTGA ATCCAGCAAT ATAAAAAGG GATTATATAC TCTGGCCAAC TGACATTCT 27867
50	CCTAGGAATG CAAAGATGGT TTAATATCCT AAAATCAATT AACATAACAT ACTATATTAA 27927
	TAAAGTATCA AACAGTATT CTCATCTTT TTTCTTTT CACAATTCT TGGTTACACT 27987
	ATCATCTCAA TAGATGCAGA AAAAGCATTG GACAAAATCC AATTCTATAAT AAAAATTCTC 28047

5	AAACTTGAAA GAGAACATCA TAAAGGCATC TATGAAAAAC CTACAGCTAA TATCATACTT 28107
	AACGATGAAA AACTGAATTA TTTTACCCCTA AGATCAAGAA TAATGCAAGC ATGTCAGCTC 28167
	TTGCAACTTC TATTCAACAT TGTACTGGAG GTTCTAGCCA GAGCAACCAT ACAATAAAA 28227
	AAAATAAAAG GCACCCAGAT TAGAAAGGAA GTCTTTATTT GCAGACAAAC TGTTTCTTTA 28287
	TGCAGAAAAC CGTCAGGAAT ACACACACAT GTTAGAACTA ATAAGTTCA CGAGGTTGCA 28347
	GGTTGCAATA TCAATATGCA AAAATACATT GAAGGCTGGG CTCAGTGGAG ATGGCATGTA 28407
	CCTTTCTGTCC CAGCTACTTG GGAGGCTGAG GTAGGAGGAT CACTTGAGGT GAGGAGTTG 28467
	AGGCTATAGT GCAATGTGAT CTTGCTGTG AATAGCCACT GCACTCGAGC CTAGGCAACA 28527
10	AAGTGAGACC CCGTCTCCAA AAAAATTTGG TATTTCTGTA TATGAACAAAT 28587
	GAATGATCTG AAAACAAGAA AATTCATTC ACCGATGGTAT TAAAAAATA AAATACAAAT 28647
	AAATTTAGCA AAATAATTAT AAAACTGTGTA CATCGAAAAT TTCAAAGCAC TCTGAGGGAA 28707
	ATTAAGATG ATCTAAATAA TTGGAGAGAC ACTCTATGAT CACTGATTGG AAAATTCAATT 28767
	CAATATTGTT AAGATAACAA TTGTCCCCAA ATTGATGCAAT GCATTCAATT TAGTCTTCAT 28827
	CAAAATTCCA GCAGGGTTTT TGCAGAAATT GACAAGCTGT ACCCAAAATG TATATGGAAA 28887
15	TGAAAAGACC CAGAAGAGCA AATAATTTT TAAAAACAAA GTTGGAAAAC TTTTACTTCC 28947
	TAATTTAAA ACTTACTATA AACCTAAAGT TATCAAGACC ATTTAGT 28994.

14. The genomic DNA of claim 1, which is derived from human.

20 15. The genomic DNA of claim 1, which is inserted into an autonomously replicable vector.

16. A transformant derived from a mammalian cell, which contains the genomic DNA claim 1.

25 17. The transformant of claim 16, which is derived from a cell selected from the group consisting of epithelial, interstitial and hemopoietic cells from mammal.

30 18. A process for preparing a polypeptide, which comprises (a) artificially expressing the DNA of claim 1, and (b) collecting a polypeptide capable of inducing the production of interferon- γ by immunocompetent cells from the resultant mixture.

19. The process of claim 18, wherein the artificial expression of the step (a) comprises a step of culturing the transformant of claim 16.

35 20. The process of claim 18, wherein the resultant mixture of the step (b) contains a culture of the transformant of claim 16.

21. The process of claim 18, wherein the polypeptide is collected by one or more techniques selected from the group consisting of salting out, dialysis, filtration, concentration, preparatory sedimentation, ion-exchange chromatography, gel filtration chromatography, adsorption chromatography, isoelectric point chromatography, hydrophobic chromatography, reversed phase chromatography, affinity chromatography, gel electrophoresis and isoelectric focusing.

40 22. The process of claim 18, wherein the polypeptide is collected by an immunoaffinity chromatography with a monoclonal antibody.

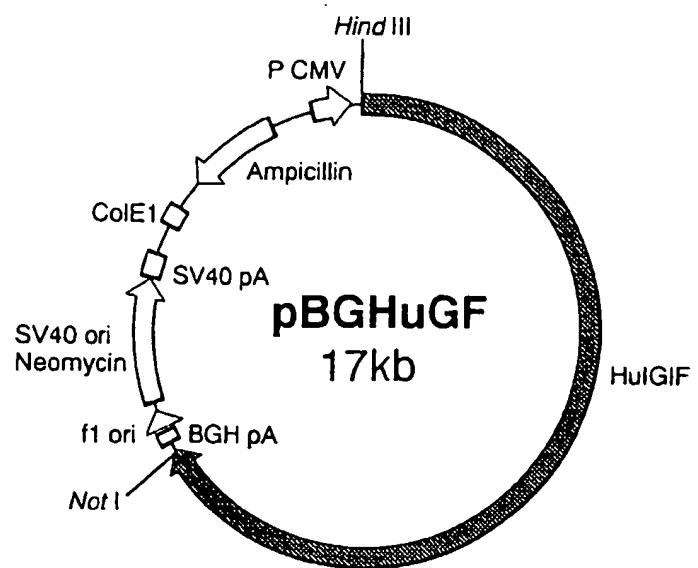


FIG.1

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.